Scaling Limit for the Diffusion Exit Problem

Sergio Angel Almada Monter

October 12, 2011

List of Figures

Contents

Chapter 1

Introduction

In this thesis we study the so called exit problem [34, Section 4.3] for small noise diffusion. This model belongs to the more general area of random perturbations of dynamical systems, which has been a very active area of research over the last 30 years [13], [34], [52]. The small noise diffusion framework has attracted the interest of both the pure and applied mathematics communities. From the mathematical standpoint it is interesting because this area has strong interactions with other important branches of mathematics such as probability theory, dynamical systems, or PDE. As regards applied mathematics, the set of problems relating to small noise diffusion has found applications in climate modeling [10], [11], electrical engineering [16], [65], [66], finance [27], [31], neural dynamics [55], [56] among others [23]. The main focus of the thesis, the exit problem, was originally motivated by applications on the reaction rate theory of chemical physics [36]. Moreover, the work presented here, although purely theoretical, was motivated by neural dynamics [55]. We proceed to describe the setting, in order to provide a more extensive background.

The setting of this problem is as follows. Given a smooth vector field $b : \mathbb{R}^d \to \mathbb{R}^d$ consider the Itô equation driven by the d-dimensional standard Wiener process W:

$$\begin{aligned} dX_\epsilon(t) &= b(X_\epsilon(t))dt + \epsilon\sigma(X_\epsilon(t))dW(t), \\ X_\epsilon(0) &= x_0. \end{aligned} \tag{1.1}$$

Here $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a smooth uniformly non-degenerate matrix valued function. That is, the matrix $a = \sigma^T\sigma$ is uniformly positive definite. Under these assumptions we can ensure that equation (1.1) has a unique strong solution (see [41] or [54] for all stochastic analysis references).

Given an initial condition $x_0 \in \mathbb{R}^d$ (or a set of initial conditions), the goal is to characterize some asymptotic properties of X_ϵ as $\epsilon \to 0$. In particular we focus on the exit from a domain problem or exit problem for short. Consider a domain (open, bounded and connected) $D \subset \mathbb{R}^d$ with piecewise smooth boundary (at least C^2). The exit problem is the study of the time

$$\tau_\epsilon^D(x) = \inf\{t > 0 : X_\epsilon(t) \in \partial D\},$$

at which X_ϵ exits, and the exit distribution $\mathbf{P}_{x_0}\{X_\epsilon(\tau_\epsilon^D) \in \cdot\}$. In this work we aim for a joint asymptotic result on the distribution of $(\tau_\epsilon^D, X_\epsilon(\tau_\epsilon^D))$ under certain assumptions for b.

As we said before, this problem had its origin in chemical physics: it is a glorified model for the speed at which chemical reactions take place. The first model of this kind was proposed by Kramers [45], we refer to [14] for a modern treatment.

From a pure mathematics perspective, the problem became of interest because it provides a framework to compute asymptotic (as $\epsilon \to 0$) properties for solutions of the Dirichlet problem

$$\nabla u_\epsilon(x) \cdot b(x) + \frac{\epsilon^2}{2}\Delta u_\epsilon(x) = 0, \quad x \in D, \tag{1.2}$$

$$u_\epsilon(x) = g(x), \quad x \in \partial D. \tag{1.3}$$

Indeed, using a relaxed version of the Feynman-Kac theorem [41, Theorem 4.4.2], the solution to this PDE can be written as the average $u_\epsilon(x) = \mathbf{E}_x g(X_\epsilon(\tau_\epsilon^D))$, where X_ϵ is the solution of (1.1) with $\sigma = \mathrm{Id}$. Solutions to other PDE's can be written as a similar average, but for this discussion we choose (1.2) since it is representative of the area. Although the asymptotic study of the function u_ϵ is now known to be strongly related to the solution to the exit problem, the first studies relied only on analytic non-probabilistic arguments. For example, in [25], [39], [40] under some assumptions on the drift b, they were able to rigorously write the solution as a formal series in ϵ. In [60] it is proved, under the assumption that $b(0) = 0$ and that D is contained in the basin of attraction of 0, that as $\epsilon \to 0$, u_ϵ converges to a constant. This result is of great importance since it means that the system forgets its initial condition. However, this is not always true. For instance in [2] using functional analysis and dynamical systems for the Levinson case (see section 1.1) it is proved that the limiting function at x depends on the orbit of x under the action of the flow generated by b. In [43] and [44] a combined approach involving probabilistic and purely analytic arguments was put into practice. Very general limit theorems were obtained, but strong restrictions on the non-linearity of the drift b were imposed.

The standard mindset in tackling this problem from the probabilistic point of view is to think of the SDE that defines X_ϵ as a (random) singular perturbation of the system $\dot{x} = b(x)$. In this context it is natural to expect that the methods used to study the exit problem lie in the intersection between probability theory and dynamical systems. Freidlin and Wentzell [34], [52] were the ones who put together a general theory in this direction. They based their theory on the Large Deviation principle for X_ϵ. This result was then used as the building block in constructing what today is known as the Freidlin-Wentzell theory. The core of this theory strongly relates the exit behavior of X_ϵ to the properties of the vector field b by providing two elements:

1. It defines a function $V : D \times \partial D \to [0, \infty]$, known as quasi-potential, that characterizes the exit distribution. Indeed, the exit distribution of X_ϵ is asymptotically concentrated on the set $V^*_{X_\epsilon(0)}$ of minimizers of $V(X_\epsilon(0), \cdot)$. Moreover, under the assumption that σ is uniformly nondegenerate, V, and hence $V^*_{X_\epsilon(0)}$, depends mostly on b. For example it can be shown that in the case of $b = \nabla\phi$, V is proportional to ϕ. This is the reason why the V function is called quasi-potential.

2. It shows that in the case in which the domain is contained in the basin of attraction of an equilibrium , $\epsilon^{-2} \log \tau_\epsilon^D$ converges in probability to the minimum of $V(X_\epsilon(0), \cdot)$. This result is of vital importance since it gives a hierarchy of transition for the case in which there are several equilibria. See [34, Section 6.5] for more background on this particular direction.

The theory came to light with a series of papers beginning with [62] and [63] until the Russian edition of the book [34] appeared. See [32] and [33] for a modern version of the theory, and [17], [19] and references therein for a stochastic partial differential equations version of the theory. In Section 1.1 we give a brief review of the Freidlin-Wentzell theory necessary to understand the motivation of this work.

In contrast with the Freidlin-Wentzell theory, that mostly relies on the large deviation principle, a modern trend relying on a path-wise approach has emerged in the last years. As a consequence, more detailed phenomena can be captured. That is the case, for example, in [4] in which a heteroclinic network is considered or in [12] in which a bifurcation problem is studied. The monograph [13] contains several examples in this direction together with applications.

In this thesis, a modern and more complete treatment of two cases is developed: the saddle case in which the vector field has a unique saddle point and the Levinson case in which the deterministic dynamics escape from the domain in a finite time. The two results when combined complete the treatment of the case in which the underlying dynamics admit a heteroclinic network as studied in [4]. We also provide a 1-dimensional example that explains how to obtain a correction to the exit time for a diffusion conditioned to exit through an unlikely exit point. The approach presented here relies heavily on the underlying dynamical structure, and combines techniques from differential equations, bifurcation theory and martingale theory.

The rest of this chapter is organized as follows. In Section 1.1 we give a brief introduction to Freidlin-Wentzell theory. In Section 1.2 we study the exit problem in the case the system $\dot{x} = b(x)$ has a saddle point. In Section 1.3 we study the escape when it takes a finite time in the so called Levinson setting. Applications of these two cases are presented in Sections 1.4.1 and 1.4.2. The general setting and a brief description of the chapters is given in Section 1.5.

1.1 Background and Motivation

In this section we gather the background tools that will allow us to explain where our results stand with respect to the classical Freidlin-Wentzell theory.

Consider X_ϵ, the strong solution to the SDE (1.1). If seen as a random perturbation, the equation for X_ϵ suggests that the process should behave like the flow generated by b:

$$\frac{d}{dt}S^t x = b(S^t x), \quad S^t x = x. \tag{1.4}$$

Indeed, through a standard martingale argument, it is easy to see that for any $\delta > 0$ there are constants $C^{(1)}_{T,\delta}$ and $C^{(2)}_{T,\delta}$ such that

$$\sup_{x\in\mathbb{R}^d} \mathbf{P}_x \left\{ \sup_{t\leq T} |X_\epsilon(t) - S^t x| > \delta \right\} \leq C^{(1)}_{T,\delta} e^{-C^{(2)}_{T,\delta}\epsilon^{-2}}. \tag{1.5}$$

This inequality can be used to show that $\mathbf{P}^\epsilon_x$, the law of X_ϵ on $C([0,T];\mathbb{R}^d)$ conditioned to $X_\epsilon(0) = x$, converges weakly (on the space $C([0,T];\mathbb{R}^d)$) to the measure concentrated on the orbit of x. See [15] for a series expansion in ϵ and [18] for a series expansion of more general stochastic flows.

The question now is to find the optimal constant $C^{(2)}_{T,\delta}$ in (1.5), or more generally to find a large deviation principle [23], [24] or Appendix A:

Theorem 1 (Freidlin-Wentzell [34]) *Let $H^1_{0,T}$ be the space of all absolutely continuous functions from $[0,T]$ to $\mathbb{R}^d$ with square integrable derivatives. Define the functional I^x_T by*

$$I^x_T(\varphi) = \frac{1}{2}\int_0^T \langle \dot{\varphi}\,(s) - b(\varphi(s)), a^{-1}(\varphi(s))(\dot{\varphi}\,(s) - b(\varphi(s)))\rangle ds, \tag{1.6}$$

if $\varphi \in H^1_{0,T}$ and $\varphi(0) = x$, and ∞ otherwise. Here b is the drift in (1.1) *and $a = \sigma^T\sigma$, with σ the diffusion matrix in* (1.1).

Then for each $x \in \mathbb{R}^d$ and $T > 0$ the family $(\mathbf{P}^\epsilon_x)_{\epsilon>0}$ satisfies a Large Deviation Principle on $C([0,T];\mathbb{R}^d)$ equipped with uniform norm at rate ϵ^2 with good rate function I^x_T.

See [9] and [51] for more large deviations results. In order to have this thesis as self contained as possible, we give a large deviation overview on Appendix A .

Informally, Theorem 1 says that if $A \subset C([0,T];\mathbb{R}^d)$ then

$$\mathbf{P}_x\left\{X_\epsilon \in A\right\} \asymp e^{-\epsilon^{-2}\inf_{\varphi\in A} I^x_T(\varphi)}.$$

Intuitively, due to (1.6), this result suggest that I^x_T serves as a measure on how costly (in terms of probability) is for the system X_ϵ not to follow the deterministic trajectory. This interpretation is essential when solving problems that require non-compact time frames, in particular, when studying the exit problem described above.

Regarding I^x_T as a cost function, it make sense to introduce

$$V(x,y) = \inf_{T>0}\{I^x_T(\varphi) : \varphi(T) = y, \varphi([0,T]) \subset D \cup \partial D\} \tag{1.7}$$

as the cost to go from x to y inside D. The function $V : D \times \partial D \to [0,\infty]$, known as the quasipotential, plays an important role on the exit problem we described:

Theorem 2 (Freidlin-Wentzell [34]) *Suppose $X_\epsilon(0) = x_0$ and let*

$$z = \inf_{y\in\partial D} V(x_0, y).$$

Then for every closed set $N \subset \partial D$ such that $\inf_{y\in N} V(x_0,y) > z$,

$$\lim_{\epsilon\to 0}\mathbf{P}_{x_0}\{X_\epsilon(\tau_\epsilon) \in N\} = 0.$$

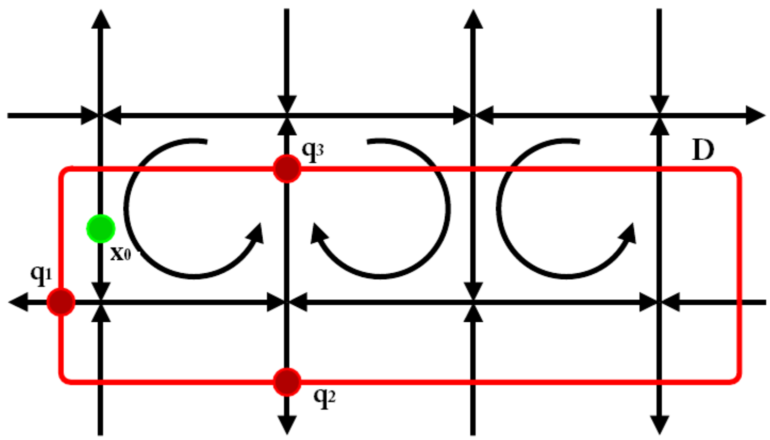

Figure 1.1: Example with several minimizers

The necessary observation derived from this theorem is that, in the limit, the exit occurs in a neighborhood of the set of minimizers of the quasipotential. The limitation is that when there are several minimizers, the result doesn't provide any distinction between them. For example, suppose the phase portrait of S is as in Figure 1.1 and D is a rectangular region as in Figure 1.1. Then the set of minimizers consists of 3 points: q_1, q_2 and q_3. Freidlin-Wentzell theory ensures that, asymptotically, the exit occurs outside this set with exponentially small probability, but doesn't distinguish between the minimizer. For example, the theory is not able to establish if it is more likely to exit on a neighborhood of q_1 or in a neighborhood of q_2. Bakhtin [4] started a theory that will allow us to compute the probability of exiting close to each of the minimizers. To complete this theory is part of the motivation for this work.

This work (see also [1]) drops some technical assumptions needed in [4] in the planar case. The program makes intensive use of normal form theory and provides small noise estimates for non-linear diffusion. With this work, the asymmetric behavior found in [4] is extended to arbitrary Hamiltonian systems on the plane. Moreover, the present text also (see [2]) provides exact scaling corrections when the flow S exits the domain in a finite time. This is a step towards to both Freidlin-Wentzell theory and Bakhtin's heteroclinic result, since extension to the case in which S is asymptotically stable can be carried out by time reversing. In this direction, a 1 dimensional example is presented in Chapter 3. This is an open and promising future research area, since it may provide scaling limits for exit points under very general

assumptions on b.

1.2 Escape from a Saddle

In this section, we assume that the system has a unique critical point and that point is a saddle. Without loss of generality, suppose that the critical point is the origin; that is, we are assuming that $0 \in \mathbb{R}^d$ is the only point $x \in D \cup \partial D$ such that $b(x) = 0$, and the matrix $A = \nabla b(0)$ has spectrum bounded away from zero with at least one eigenvalue with positive real part and one eigenvalue with negative real part. In other words, there is a pair of integers $\nu, \mu \geq 1$ such that the eigenvalues $\lambda_1, ..., \lambda_d$ of A satisfy

$$\operatorname{Re}\lambda_1 = ... = \operatorname{Re}\lambda_\nu > \operatorname{Re}\lambda_{\nu+1} \geq ... \geq \operatorname{Re}\lambda_\mu > 0 > \operatorname{Re}\lambda_{\mu+1} \geq ... \geq \operatorname{Re}\lambda_d.$$

Under this assumptions, it is well known [64, Theorem 3.2.1] that $\bar{D} = D \cup \partial D$ can be decomposed as $\bar{D} = \{0\} \cup \mathcal{W}^u \cup \mathcal{W}^c \cup \mathcal{W}^s$, where

$$\mathcal{W}^u = \{x \in \bar{D} : \lim_{t\to-\infty} S^t x = 0, \text{ and for some } s \geq 0, S^{(-\infty,s)}x \subset D \text{ and } S^{(s,\infty)}x \cap \bar{D} = \emptyset\},$$

$$\mathcal{W}^c = \{x \in \bar{D} : S^{(s_1,s_2)}x \subset D \text{ and } S^{[s_1,s_2]^c}x \cap \bar{D} = \emptyset, s_1 \leq 0 \text{ and } s_2 \geq 0\},$$

and,

$$\mathcal{W}^s = \{x \in \bar{D} : \lim_{t\to\infty} S^t x = 0, \text{ and for some } s \leq 0, S^{(-\infty,s)}x \cap \bar{D} = \emptyset \text{ and } S^{(s,\infty)}x \subset D\}.$$

Here, for an interval $A \subset \mathbb{R}$, $S^A x$ denotes the set

$$S^A x = \{S^t x : t \in A\}.$$

We are ready to state the theorem in [44] concerning the exit time τ_ϵ:

Theorem 3 *If $x \in D \cap \mathcal{W}^s$, then*

$$-\frac{\tau_\epsilon}{\log \epsilon} \xrightarrow{\mathbf{P}} \frac{1}{\operatorname{Re}\lambda_1}, \quad \epsilon \to 0.$$

Consider the (deterministic) time

$$T(x) = \inf\{t > 0 : S^t x \in \partial D\},$$

then, if $x \in (\mathcal{W}^c \cup \mathcal{W}^u) \cap D$,

$$\tau_\epsilon \xrightarrow{\mathbf{P}} T(x), \quad \epsilon \to 0.$$

In order to state the corresponding theorem for the exit distribution, denote $\Gamma_{\max}$ the generalized eigenspace of A which corresponds to $\lambda_1, ..., \lambda_\nu$. The Hadamard–Perron theorem [53, Section 2.7], [64, Theorem 3.2.1] states that there is a ν-dimensional S^t-invariant submanifold W_{max} tangent to $\Gamma_{\max}$ at the origin. Note that the intersection $Q_{\max} = W_{\max} \cap \partial D$ is not empty. Moreover, in the case of $\nu > 1$, $Q_{\max}$ is a $\nu - 1$-dimensional manifold, while for $\nu = 1$ consists of two points: $Q_{\max} = \{q_-, q_+\}$. The result in [44] reads:

Theorem 4 *If $x \in D \cap \mathcal{W}^s$ and $\nu > 1$, then for any relatively open subset $G \subset \partial D$ such that $Q_{\max} \subset G$, it holds that*

$$\lim_{\epsilon \to 0} \mathbf{P}_x\{X_\epsilon(\tau_\epsilon) \in G\} = 1.$$

If $\nu = 1$ then the measure $\mathbf{P}_\epsilon^x(\cdot) = \mathbf{P}\{X_\epsilon(\tau_\epsilon) \in \cdot | X_\epsilon(0) = x\}$ converges weakly to the measure $\frac{1}{2}\delta_{q_-} + \frac{1}{2}\delta_{q_+}$, where δ_z is the probability measure concentrated at z.

If $x \in (\mathcal{W}^c \cup \mathcal{W}^u) \cap D$, $\mathbf{P}_\epsilon^x(\cdot)$ converges weakly to the measure $\delta_{S^{T(x)}x}$.

In the work by Day [22] a refinement to the theorem about the exit time is given in the 2−dimensional situation. He proved that $\lambda_1 \tau_\epsilon$ can be written as a sum between $-\log \epsilon$ and a tight correction. He gave a precise asymptotic description for the distribution of the correction:

Theorem 5 *If $d = 2$ and $X_\epsilon(0) \in \mathcal{W}^s \cap D$, then*

$$\lambda_1 \tau_\epsilon^D + \log \epsilon \to \mathcal{K} + C_\nu$$

in distribution. Here $\mathcal{K}$ and C_ν are independent random variables. Moreover, $\mathcal{K}$ has a density with respect to Lebesgue measure given by

$$d\mathcal{K} = \frac{2}{\sqrt{\pi}} e^{-(x + e^{-2x})} dx,$$

and C_ν is a Bernoulli random variable with $\mathbf{P}\{C_\nu = C_\pm\} = 1/2$, where $C_\pm$ are a constants depending only on b and σ.

This theorem complements previous work by Mikami [49] which established the decay in the distribution of $-\frac{\lambda_1}{\log \epsilon} \tau_\epsilon^D$:

Theorem 6 *For an arbitrary $d \geq 1$ and $X_\epsilon(0) \in \mathcal{W}^s \cap D$ for $T \in (0, 1)$ it holds that*

$$\lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \log \left(-\log \mathbf{P} \left\{ -\frac{\lambda_1}{\log \epsilon} \tau_\epsilon^D < T \right\} \right) = T - 1,$$

while for $T > 1$,

$$\lim_{\epsilon \to 0} \frac{1}{\log \epsilon} \mathbf{P} \left\{ -\frac{\lambda_1}{\log \epsilon} \tau_\epsilon^D > T \right\} = (T-1)/2.$$

Theorem 6 is a good refinement of Theorem 3, but it doesn't prove that the distribution of the difference $\lambda_1 \tau_\epsilon^D + \log \epsilon$ is tight as in Theorem 5. The tails established in Theorem 6 are consistent with the tails of Theorem 5.

Bakhtin [5] gave a refinement of the last theorems. He proved that in any dimension the family of random variables $\tau_\epsilon + \lambda_1^{-1} \log \epsilon$ converges in distribution and he identified the limit. He also improved the result about the convergence on the exit distribution in Theorem 4. In the case $\nu = 1$, he showed that the factor $1/2$ in the limiting distribution $\frac{1}{2}\delta_{q_+} + \frac{1}{2}\delta_{q_-}$ comes from the symmetry of a certain Gaussian random variable.

Theorem 7 *In the case $\nu = 1$, parametrize the manifold $W_{\max}$ as a C^1-curve γ that can be locally represented as a graph over the (one dimensional) space $\Gamma_{\max}$. Let the times $t(\pm\delta, q_\pm)$ be the time it takes to the deterministic flow to go from $\gamma(\pm\delta)$ to $q_\pm$:*

$$t(\pm\delta, q_\pm) = T(\gamma(\pm\delta)),$$

with $T(x)$ defined as in Theorem 3.

Then, the numbers

$$h_\pm = \lim_{\delta \to 0} \left(\frac{\log \delta}{\lambda_1} + t(\pm\delta, q_\pm) \right)$$

are well defined and, in the case where $\sigma = I$, and $x \in D \cap \mathcal{W}^s$, there is a Gaussian random variable $\mathcal{N}$ and a number $\kappa = \kappa(x) > 0$ such that :

1. *As $\epsilon \to 0$ the convergence*

$$X_\epsilon(\tau_\epsilon) \xrightarrow{\mathbf{P}} q_+ \mathbf{1}_{\{\mathcal{N}>0\}} + q_- \mathbf{1}_{\{\mathcal{N}<0\}},$$

and

$$\tau_\epsilon + \frac{1}{\lambda_1} \log \epsilon \xrightarrow{\mathbf{P}} h_+ \mathbf{1}_{\{\mathcal{N}>0\}} + h_- \mathbf{1}_{\{\mathcal{N}<0\}} - \frac{1}{\lambda_1} \log(\kappa \mathcal{N})$$

hold in probability.

2. *As $\epsilon \to 0$ the distribution of the random vector $(X_\epsilon(\tau_\epsilon), \tau_\epsilon + \frac{1}{\lambda_1}\log\epsilon)$ converges weakly to the measure*

$$\frac{1}{2}\delta_{q_+} \times \mu_{h_+,\omega} + \frac{1}{2}\delta_{q_-} \times \mu_{h_-,\omega},$$

where $\mu_{h_\pm,\omega}$ is the distribution of

$$h_\pm - \frac{1}{\lambda_1}\log(\kappa\mathcal{N}).$$

The proof of Theorems 5 and 7 is based on the study of the linear system and then an approximation to the non-linear one. The steps followed for a prototypical 2−dimensional system is presented in the next section. Part of this thesis focus is on adapting this methodology to the non-linear case.

1.2.1 A 2-dimensional linear example

For a fixed $\delta > 0$, consider the domain $D = (-\delta, \delta) \times (-\delta, \delta) \subset \mathbb{R}^2$. Given two positive numbers $\lambda_\pm > 0$, we sketch the solution to the exit problem from D for the diffusion $X_\epsilon = (x^1_\epsilon, x^2_\epsilon)$ given by

$$\begin{aligned} dX_\epsilon(t) &= \mathrm{diag}(\lambda_+, -\lambda_-)X_\epsilon(t)dt + \epsilon dW(t), \\ X_\epsilon(0) &= (0, x_0) \in D. \end{aligned}$$

Here, for a column vector $v = (v_1, v_2)$, v_1 is the first coordinate and v_2 is the second one.

Using Itô's formula [41, Theorem 3.3.3] in each coordinate we write the Duhamel principle for x^1_ϵ and x^2_ϵ as

$$x^1_\epsilon(t) = \epsilon e^{\lambda_+ t}\int_0^t e^{-\lambda_+ s}dW_1(s), \tag{1.8}$$

$$x^2_\epsilon(t) = e^{-\lambda_- t}x_0 + \epsilon\int_0^t e^{-\lambda_-(t-s)}dW_2(s). \tag{1.9}$$

These two identities are the main ingredient in this development.

We will show that X_ϵ exits D along $(\delta, 0)$ or $(-\delta, 0)$. Consider the time at which x^1_ϵ exits the interval $(-\delta, \delta)$:

$$\tau^\delta_\epsilon = \inf\{t > 0 : |x^1_\epsilon(t)| = \delta\}.$$

The program is to compute τ^δ_ϵ based on the path-wise properties of X_ϵ. Use the identity obtained for τ^δ_ϵ to characterize $X_\epsilon(\tau^\delta_\epsilon)$ from which we will deduce

that $\mathbf{P}\{\tau_\epsilon^\delta = \tau_\epsilon^D\} \to 1$ as $\epsilon \to 0$. Hence, we can obtain the limiting behavior of $(\tau_\epsilon^D, X_\epsilon)$ based on the identities we have obtained for $(\tau_\epsilon^\delta, X_\epsilon(\tau_\epsilon^\delta))$. This will establish a result in the spirit of Theorem 7.

First, note that $\tau_\epsilon < \infty$ with probability 1. This is a classical fact that we prove in Appendix B for completeness. On the set $\{\tau_\epsilon^\delta < \infty\}$, define the random variable $\mathcal{N}_\epsilon$ by

$$\mathcal{N}_\epsilon = \int_0^{\tau_\epsilon^\delta} e^{-\lambda_+ s} dW_1(s).$$

An application of Duhamel's principle (1.8) for x_ϵ^1 and the definition of τ_ϵ^δ establishes the equality $\delta = \epsilon e^{\lambda_+ \tau_\epsilon^\delta} |\mathcal{N}_\epsilon|$ with probability 1. This identity implies that, with probability 1,

$$\tau_\epsilon^\delta = -\frac{1}{\lambda_+} \log \epsilon + \frac{1}{\lambda_+} \log \left(\frac{\delta}{\mathcal{N}_\epsilon} \right). \tag{1.10}$$

Using equality (1.10) together with (1.8) and (1.9) it holds that

$$x_\epsilon^1(\tau_\epsilon^\delta) = \delta \operatorname{sgn}(\mathcal{N}_\epsilon), \tag{1.11}$$

and

$$x_\epsilon^2(\tau_\epsilon^\delta) = \epsilon^{\lambda_-/\lambda_+} x_0 \left(\frac{|\mathcal{N}_\epsilon|}{\delta} \right)^{\lambda_-/\lambda_+} + \epsilon \int_0^{\tau_\epsilon^\delta} e^{-\lambda_-(t-s)} dW_2(s), \tag{1.12}$$

both with probability 1. Hence, if we can establish tightness for the distribution of the family of random variables $(\mathcal{N}_\epsilon)_{\epsilon>0}$, the fact that $\mathbf{P}\{\tau_\epsilon^\delta = \tau_\epsilon^D\}$ converges to 1 would be a consequence of (1.11), (1.12) and the tightness of the distribution of the stochastic integral in (1.12).

In order to get the tightness result, we need to analyze the time τ_ϵ^δ without any reference to (1.10). We can show two properties (see Appendix B for their proofs):

1. For every $\delta > 0$, $\tau_\epsilon^\delta \to \infty$ in probability as $\epsilon \to 0$.

2. As a consequence to the last point, as $\epsilon \to 0$,

$$\mathcal{N}_\epsilon \xrightarrow{\mathbf{P}} \int_0^\infty e^{-\lambda_+ s} dW_1(s).$$

Let $\mathcal{N}$ be the limit Gaussian random variable of $(\mathcal{N}_\epsilon)_{\epsilon>0}$ in the second observation above. Then, we have proved the following lemma:

Lemma 8

$$\tau_\epsilon^\delta + \frac{1}{\lambda_+}\log\epsilon \xrightarrow{\mathbf{P}} \frac{1}{\lambda_+}\log\left(\frac{\delta}{|\mathcal{N}|}\right), \quad \epsilon \to 0.$$

We apply this lemma to the exit distribution of X_ϵ. Before that, let us denote $\mathcal{N}_-$ a zero mean Gaussian random variable with variance $(2\lambda_-)^{-1}$ independent of $\mathcal{N}$. It is possible to prove that

$$\int_0^{\tau_\epsilon^D} e^{-\lambda_-(t-s)}dW(s) \to \mathcal{N}_-, \quad \epsilon \to 0,$$

in distribution. This convergence combined with the convergence of τ_ϵ^δ and $\mathcal{N}_\epsilon$ used in (1.11) and (1.12) implies that on the set $\{\tau_\epsilon^D = \tau_\epsilon^\delta\}$,

$$X_\epsilon(\tau_\epsilon^D) = \delta(\operatorname{sgn}\mathcal{N}_\epsilon, 0) + \epsilon^{(\lambda_-/\lambda_+)\wedge 1}(0, \xi_\epsilon). \tag{1.13}$$

Here $(\xi_\epsilon)_{\epsilon>0}$ is a family of random variables that satisfies

$$\xi_\epsilon \to \left(\frac{|\mathcal{N}|}{\delta}\right)^{\lambda_-/\lambda_+} x_0\mathbf{1}_{\{\lambda_-\le\lambda_+\}} + \mathcal{N}_-\mathbf{1}_{\{\lambda_-\ge\lambda_+\}}, \quad \epsilon \to 0, \tag{1.14}$$

in distribution. Moreover, it can be shown that when $\lambda_- < \lambda_+$ this convergence holds in probability.

Hence Theorem 7 holds with $\gamma(t) = (t, 0)$ up to time re-parametrization.

1.2.2 Analysis and generalization

The simplified argument of last section not only recovers, for this simple linear case, Theorem 7 but also provides more information. From (1.12) we can see that when $\lambda_- \le \lambda_+$ the exit distribution has a bias in the stable direction. In this case, the second coordinate of the exit distribution is not centered. Indeed, from (1.13) and (1.14) we can identify 3 cases:

1. When $\lambda_- > \lambda_+$, the second coordinate of X_ϵ converges to $\mathcal{N}_-$. Hence, the exit has a centered distribution in the stable direction. We refer to this case as the symmetric case and is illustrated in Figure 1.2.

2. When $\lambda_- = \lambda_+$, the second coordinate of X_ϵ converges to $\left(\frac{|\mathcal{N}|}{\delta}\right)^{\lambda_-/\lambda_+} x_0 + \mathcal{N}_-$. Hence, the exit has a bias in the stable direction due to the initial condition x_0. We refer to this case as the asymmetric case.

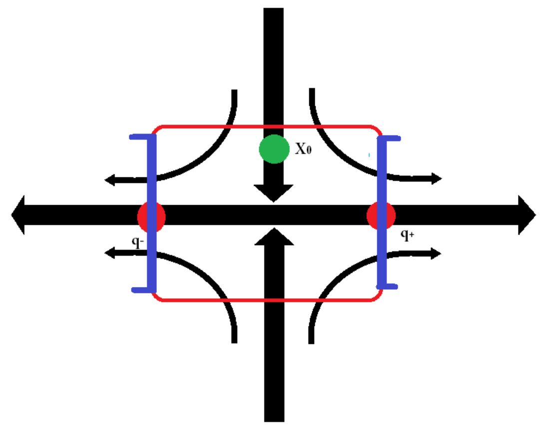

Figure 1.2: Symmetric Case.

3. When $\lambda_- < \lambda_+$, the second coordinate of X_ϵ converges to $\left(\frac{|\mathcal{N}|}{\delta}\right)^{\lambda_-/\lambda_+} x_0$. Hence, the exit has a very strong bias in the stable direction due to the initial condition x_0. We refer to this case as the strongly asymmetric case and is illustrated in Figure 1.3

The consequences of such an asymmetry, when there is one, have been explored in [4], and it turned out to be an important improvement to Freidlin-Wentzell theory for a particular situation. We will summarize this improvement in Section 1.4.2. For now let us comment about how general the argument of last Section 1.2.1 really is.

The immediate limitation of the argument presented in Section 1.2.1 is that it is mostly based on explicit representations for the solution of X_ϵ. In [5], [4] and [22] a linear approximation to the original process X_ϵ is made. The non-optimal feature of this procedure is that we lose all the identities, and we have just approximations. This is not acceptable if we are interested on computing the properties of the aforementioned asymmetry.

A similar but structurally different argument is presented here in Chapter 2 (see [4] also) in which a change of variable is introduced to linearize the system locally. Consider the process $Y_\epsilon(t) = f(X_\epsilon(t))$ and for the moment assume f to be smooth. Then, the new process Y_ϵ solves the SDE:

$$dY_\epsilon = \left(\nabla f^{-1}(Y_\epsilon)\right)^{-1} b(f^{-1}(Y_\epsilon))dt + \epsilon\tilde{\sigma}(Y_\epsilon)dW + \epsilon^2\Psi(Y_\epsilon)dt,$$

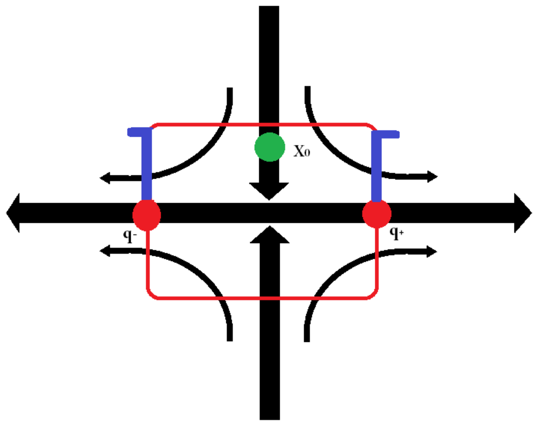

Figure 1.3: Strongly Asymmetric Case.

for some smooth (depending on the function σ) functions $\Psi : \mathbb{R}^d \to \mathbb{R}^d$ and $\tilde{\sigma} : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$. If we can choose the function f to be such that

$$\left(\nabla f^{-1}(y)\right)^{-1} b(f^{-1}(y)) = Ay, \tag{1.15}$$

then the proof of the theorem will follow almost identically as the case in Section 1.2.1 (only with multiplicative noise instead of additive one) giving the asymmetry for the general case. The existence of such a transformation belongs to the study of conjugation in Ordinary Differential Equations. The main result of the latter theory (the so called Hartman-Grobman Theorem [37, Theorem IX.7.1], [53, Section 2.8]) guarantees the existence of a homomorphism f that solves (1.15). This is not enough for our purposes since we need f to be C^2 in order to use Itô's formula. In [4] is assumed that the transformation f exist and is C^2. In order to become aware on how restrictive (if restrictive at all) is this hypothesis, we need to study transformations f that satisfy a relaxed (in a sense explained below) version of (1.15). Such transformations are the main subject of study in normal form Theory [20], [42], [53], [64].

We will summarize the main ideas in normal form theory. Following [38], we call a set of complex numbers $\lambda = (\lambda_1, \lambda_2, ..., \lambda_d)$ non-resonant if there are no integral relations between them of the form $\lambda_j = \alpha \cdot \lambda$, where $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{Z}_+^d$ is a multi-index with $|\alpha| = \alpha_1 + ... + \alpha_d \geq 2$. Otherwise,

we say that it is resonant. Moreover, a resonant λ is said to be one-resonant if all the resonance relations for λ follow from a single resonance relation. A monomial $x^\alpha e_j = x_1^{\alpha_1}...x_d^{\alpha_d}e_j$ is called a resonant monomial of order R if $\alpha\cdot\lambda = \lambda_j$ and $|\alpha| = R$. Normal form theory asserts (see [20], [38]) that for any pair of integers $R \geq 1$ and $k \geq 1$, there are two neighborhoods of the origin Ω_f and Ω_g and a C^k-diffeomorphism $f : \Omega_f \to \Omega_g$ with inverse $g : \Omega_g \to \Omega_f$ such that

$$\left(\nabla f^{-1}(y)\right)^{-1} b(f^{-1}(y)) = Ay + P(y) + \mathcal{R}(y), \quad y \in \Omega_g \tag{1.16}$$

where P is a polynomial containing only resonant monomials of order at most R and $\mathcal{R}(\zeta) = O(|\zeta|^{R+1})$. If λ is non-resonant, then f can be chosen so that both P and $\mathcal{R}$ in (1.16) are identically zero. Moreover, due to [38, Theorem 3,Section 2], if λ is one-resonant then f can be chosen so that $\mathcal{R}$ in (1.16) is identically zero.

The result in [4] only includes the non-resonant case. When applied to heteroclinic networks (network of saddles interconnected to each others) this assumption impose a restriction by requiring each critical point to be non-resonant. In particular, typical Hamiltonian systems (that usually present heteroclinic structures) have resonant relations due to the symplectic structure [20].

In this work, we give a complete solution to the 2−dimensional case with no assumptions about resonance, and with random initial conditions. The theorem (see [1] also) informally reads as

Theorem 9 *Let X_ϵ be the solution of equation* (1.1) *with initial condition given by*

$$X_\epsilon(0) = x_0 + \epsilon^\alpha \xi_\epsilon,$$

where $\alpha \in (0,1)$ and $(\xi_\epsilon)_{\epsilon>0}$ is a family of random vectors that converges weakly to the random vector ξ_0. We assume that $x_0 \in \mathcal{W}^s$ and ξ_0 is such that

$$\mathbf{P}\{b(x_0)||\xi_0\} = 0,$$

where $||$ means collinearity of vectors.

Then, there is a family of random vectors $(\phi_\epsilon)_{\epsilon>0}$, a family of random variables $(\psi_\epsilon)_{\epsilon>0}$, and the number

$$\beta = \begin{cases} 1, & \alpha\lambda_- \geq \lambda_+ \\ \alpha\frac{\lambda_-}{\lambda_+}, & \alpha\lambda_- < \lambda_+ \end{cases} \tag{1.17}$$

such that

$$X_\epsilon(\tau_\epsilon^D) = q_{\mathrm{sgn}(\psi_\epsilon)} + \epsilon^\beta \phi_\epsilon.$$

The random vector

$$\Theta_\epsilon = \left(\psi_\epsilon, \phi_\epsilon, \tau_\epsilon^D + \frac{\alpha}{\lambda_+}\ln\epsilon\right)$$

converges in distribution as $\epsilon \to 0$ to the limit Θ_0 that can be identified. The exit distribution exhibits the behavior presented at the beginning of Section 1.2.2.

The proof of this theorem is divided in two steps. One, is the study of the diffusion X_ϵ when is close to the origin. The second, is the study of the diffusion X_ϵ far from the origin. Here, the meaning of close and far relies on whether or not the system can be conjugated to its Normal Form.

By the study of X_ϵ close to the critical point, we mean $X_\epsilon \in B$, where B is a neighborhood of the origin in which normal form conjugation is valid. In this case, the analysis has two parts. The first part is when the diffusion starts along the stable direction. To study this part, we study the diffusion until the the projection of X_ϵ along the stable direction dominates over the noise level. We achieve this by posing the problem as an exit problem from the strip $[-\epsilon^{\bar\alpha}, \epsilon^{\bar\alpha}] \times B$, for some $\bar\alpha \in (0, \alpha)$. For the second part we study the exit problem from B with the initial condition being the exit distribution obtained in the first part. In order for this program to be successful, we require very precise path-wise expressions for the diffusion. By using the general form of the resonances we are able to preserve the essence of the argument in Section 1.2.1.

The analysis of the system far from the origin is used twice. The first time is when the system starts along the stable manifold $\mathcal{W}^s$. The second time is when X_ϵ is about to exit the domain and it probably has some bias. Our study is based on a series expansion in powers of $\epsilon > 0$. This expansion is inspired by [15], but it requires additional geometric arguments. These results are of independent interest, so we start a new section to describe them.

1.3 Levinson Case.

This section is devoted to the Levinson case. We will first review the history of the problem and then outline our contribution.

Given an initial condition $x_0 \in D$, Levinson condition is a hypothesis associated to the flow S and the domain D. This case was originally formulated in [46] with a PDE flavor, we state the condition as presented by Freidlin [35, Chapter 2].

Condition 10 (Levinsion) *The flow S satisfies the Levinson condition at $x_0 \in D$ with respect to D if the following holds:*

1. *The exit time*
$$T(x_0) = \inf\{t > 0 : S^t x_0 \in \partial D\},$$
is finite.

2. *The flow $S^t x_0$ leaves the domain immediately after $T(x)$. That is, there is a $\delta > 0$ such that $S^{T(x_0)+s} \notin D \cup \partial D$ for all $s \in (0, \delta)$.*

We say that the domain $D \subset \mathbb{R}^d$ satisfies Levinson condition if properties 1 and 2 are satisfied at each $x \in D$.

As mentioned in the introduction, Levinson [46], [47] was originally interested in studying the behavior of the solution of the PDE:

$$\begin{aligned} \nabla u_\epsilon(x) \cdot b(x) + \frac{\epsilon^2}{2} \Delta u_\epsilon(x) = 0, &\quad x \in D, \\ u_\epsilon(x) = g(x), &\quad x \in \partial D. \end{aligned} \tag{1.18}$$

The claim is that the solution to this PDE has to converge to the solution of the unperturbed PDE:

Theorem 11 ([46]) *Under the Levinson condition 10, there is a unique (maybe generalized) solution to both, the perturbed problem* (1.18) *and the unperturbed problem*

$$\begin{aligned} \nabla u_0(x) \cdot b(x) = 0, &\quad x \in D, \\ u_0(x) = g(x), &\quad x \in \partial D. \end{aligned}$$

Let $u_0 : \mathbb{R}^d \to \mathbb{R}^d$ be a solution to the unperturbed problem. Then, if $g : \partial D \to \mathbb{R}^d$ is smooth,
$$u_\epsilon(x) = u_0(x) + \epsilon v_\epsilon(x),$$
where $v_\epsilon(x)$ is a locally bounded function for each $\epsilon > 0$.

The proof of this theorem is based on a series expansion of u_ϵ along the characteristics of (1.18). In order to prove the result with this idea, several technical challenges had to be overcome in [46]. Contrastingly, as pointed out in [35], the probabilistic approach here is simpler and cleaner. Indeed, once we know inequality (1.5), the convergence in Theorem 11 is an immediate consequence of the stochastic representation $u_\epsilon(x) = \mathbf{E}_x g(X_\epsilon(\tau_\epsilon^D))$, where

X_ϵ solves (1.1) with $\sigma = \mathrm{Id}$. To get the exact behavior a series expansion of the processes X_ϵ has to be made as in [15] and [34, Chapter 2].

In this thesis we develop a path-wise approach to this problem. We give a geometrical characterization of the exit point $X_\epsilon(\tau_\epsilon)$, and joint properties of $(X_\epsilon(\tau_\epsilon), \tau_\epsilon)$ are obtained. We start by obtaining a generalization of the series expansion given in [15]. This result serves as backbone of our proof and has independent interest on itself.

In order to present our result, we need further notation. Write b as

$$b(x) = b(y) + \nabla b(y)(x-y) + Q_1(y, x-y), \quad x, y \in \mathbb{R}^d,$$

where

$$|Q_1(u,v)| \le K|v|^2$$

for some constant $K > 0$ and any $u, v \in \mathbb{R}^d$.

Denote by $\Phi_x(t)$ the linearization of S along the orbit of x:

$$\frac{d}{dt}\Phi_x(t) = A(t)\Phi_x(t), \;\; \Phi_x(0) = I, \tag{1.19}$$

where $A(t) = \nabla b(S^t x)$ and I is the identity matrix. We can state our first lemma:

Lemma 12 *Consider the initial value problem*

$$dX_\epsilon(t) = \left(b(X_\epsilon(t)) + \epsilon^{\alpha_1}\Psi_\epsilon(X_\epsilon(t))\right)dt + \epsilon\sigma(X_\epsilon(t))dW \tag{1.20}$$

$$X_\epsilon(0) = x_0 + \epsilon^{\alpha_2}\xi_\epsilon, \quad \epsilon > 0. \tag{1.21}$$

where, for each ϵ, Ψ_ϵ is a deterministic Lipschitz vector field on $\mathbb{R}^d$ converging uniformly to a limiting Lipschitz vector field Ψ_0. Both α_1 and α_2 are positive scaling exponents. The family of random variables $(\xi_\epsilon)_{\epsilon>0}$ converges, as before, to ξ_0 in distribution as $\epsilon \to 0$.

Let

$$\begin{aligned}\phi_\epsilon(t) = {} & \epsilon^{\alpha_2-\alpha}\Phi_{x_0}(t)\xi_\epsilon + \epsilon^{\alpha_1-\alpha}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}\Psi_0(S^s x_0)ds \\ & + \epsilon^{1-\alpha}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}\sigma(S^s x_0)dW(s),\end{aligned}$$

and

$$\begin{aligned}\phi_0(t) = {} & \mathbf{1}_{\{\alpha_2=\alpha\}}\Phi_{x_0}(t)\xi_0 + \mathbf{1}_{\{\alpha_1=\alpha\}}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}\Psi_0(S^s x)ds \\ & + \mathbf{1}_{\{1=\alpha\}}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}^{-1}(s)\sigma(S^s x_0)dW(s), \quad t > 0.\end{aligned}$$

Then,

$$X_\epsilon(t) = S^t x_0 + \epsilon^\alpha \varphi_\epsilon(t)$$

holds almost surely for every $t > 0$, where $\varphi_\epsilon(t) = \phi_\epsilon(t) + r_\epsilon(t)$, and r_ϵ converges to 0 uniformly over compact time intervals in probability. Moreover, for any $T > 0$, $\phi_\epsilon \to \phi_0$, in distribution in $C[0,T]$ equipped with uniform norm.

The reason to consider (1.20) instead of the more standard (1.1) will become evident in Section 1.4.1. For now, let us observe that, although inequality 1.5 may not hold, it is still true that

$$\sup_{t \le T} |X_\epsilon(t) - S^t x_0| \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0.$$

Hence, under the Levinson Condition 10 the exit of X_ϵ from D will occur on a finite (still random) time and very close to the deterministic exit $z = S^{T(x_0)}x_0$. A better understanding of this convergence is the main result in this section.

The main theorem provides a scaling limit to the distribution of $(\tau_\epsilon, X_\epsilon(\tau_\epsilon))$. In order to understand the theorem, we regard the boundary of D as an hypersurface M in $\mathbb{R}^d$. In general, for an hypersurface M in $\mathbb{R}^d$, denote the tangent space of M at the point $z \in M$ as T_zM. Further, we denote the (algebraic) projection onto span$(b(z))$ as $\pi_b : \mathbb{R}^d \to \mathbb{R}$, and the (geometric) projection onto T_zM along span$(b(z))$ as $\pi_M : \mathbb{R}^d \to T_zM$. In other words, for any vector $v \in \mathbb{R}^d$, $\pi_b v \in \mathbb{R}$ and $\pi_M v \in T_zM$ are the unique number and vector such that

$$v = \pi_b v \cdot b(z) + \pi_M v. \tag{1.22}$$

With this notation in mind, we are ready to state the theorem:

Theorem 13 *Let M be an hypersurface in $\mathbb{R}^d$. Let X_ϵ be the solution of* (1.20) *with initial condition* (1.21). *Consider τ_ϵ and $T(x_0)$ the exit time from M of X_ϵ and S respectively. If $\alpha = \alpha_1 \wedge \alpha_2 \wedge 1$ and $z = S^{T(x_0)}x_0$, then*

$$\epsilon^{-\alpha}(\tau_\epsilon - T, X_\epsilon(\tau_\epsilon) - z) \to (-\pi_b\phi_0(T), \pi_M\phi_0(T)), \tag{1.23}$$

in distribution. Here π_b and π_M are as in (1.22).

1.4 Applications

1.4.1 Conditioned diffusions in 1 dimension

Throughout this section, we restrict ourselves to the 1-dimensional situation. In particular, let, for each $\epsilon > 0$, X_ϵ be a weak solution of the following (1

dimensional) SDE:

$$\begin{aligned} dX_\epsilon(t) &= b(X_\epsilon(t))dt + \epsilon\sigma(X_\epsilon(t))dW(t), \\ X_\epsilon(0) &= x_0, \end{aligned}$$

where b and σ are C^1 functions on $\mathbb{R}$, such that $b(x) < 0$ and $\sigma(x) \neq 0$ for all x in an interval $[a_1, a_2]$ containing x_0. We want to study the exit of such an interval $D = [a_1, a_2]$, that is

$$\tau_\epsilon = \inf\{t \geq 0 : \ X_\epsilon(t) = a_1 \text{ or } a_2\}.$$

Let $B_\epsilon = \{X_\epsilon(\tau_\epsilon) = a_2\}$, and note that, since $b < 0$, $\lim_{\epsilon\to 0} \mathbf{P}(B_\epsilon) = 0$. More precise estimates on the asymptotic behavior of $\mathbf{P}(B_\epsilon)$ can be obtained in terms of large deviations. However, our interest is to study the process X_ϵ conditioned on the rare event B_ϵ.

In this case $T(x_0)$, the time it takes for the flow S generated by $-b$ starting at x_0 to reach a_2, is given by

$$T(x_0) = -\int_{x_0}^{a_2} \frac{1}{b(x)} dx.$$

It is known from our basic Lemma 12 that $\tau_\epsilon \to T(x_0)$ as $\epsilon \to 0$ in probability. But the correction was not known so far.

The idea is to condition the diffusion to the event B_ϵ and note that this conditioned process solves a martingale problem (hence is a diffusion) and the result from Section 1.3 are applicable. Hence, we have the lemma:

Lemma 14 *Conditioned on B_ϵ, the process X_ϵ is a diffusion with the same diffusion coefficient as the unconditioned process, and with the drift coefficient given by*

$$b_\epsilon(x) = b(x) + \epsilon^2\sigma^2(x)\frac{h_\epsilon(x)}{\int_{a_1}^x h_\epsilon(y)dy},$$

where

$$h_\epsilon(x) = \exp\left\{-\frac{2}{\epsilon^2}\int_{a_1}^x \frac{b(y)}{\sigma^2(y)}dy\right\}.$$

With this lemma and the help of an analogy of Laplace's method the main theorem in this direction is:

Theorem 15 *Conditioned on B_ϵ, the distribution of $\epsilon^{-1}(\tau_\epsilon - T(x_0))$ converges weakly to a centered Gaussian distribution with variance*

$$-\int_{x_0}^{a_2} \frac{\sigma^2(y)}{b^3(y)}dy.$$

The result is of relevance not only because of the correction itself, but also, because is the first step of analysis for diffusions conditioned on rare events. Such a tool may lead to a general theory of correction in small noise systems.

1.4.2 Planar Heteroclinic Networks

A further application of our results is to the theory of Noisy Heteroclinic Networks first proposed in [4]. Our presentation applies only to the 2-dimensional situation. See [6] for a survey in this direction.

In Section 1.4.2 we give an intuitive presentation of the argument in [4]. The general theory developed in [4] is presented in Section 1.4.2. In subsection 1.4.2 relations of this result to the current text are highlighted. In this section, we also introduce the idea of a random Poincaré map.

Intuitive argument

We study the exit problem of the diffusion (1.1) from a domain D. Consider the vector field $b : \mathbb{R}^2 \to \mathbb{R}^2$ which has finite set of critical points $\mathcal{Z} = (\zeta_k)_{k=1}^N$ inside $\bar{D}$. We assume that S admits an heteroclinic network.

A heteroclinic network for the flow S is an invariant that contains at most countable number of saddles connected with each other. For simplicity, in this section we suppose that S admits an heteroclinic network with a finite set of critical points $\mathcal{Z} = (\zeta_k)_{k=1}^N$ inside $\bar{D}$. Precisely, we assume the following:

1. Each critical point ζ_k is a saddle point of the flow S. That is, $b(\zeta_k) = 0$ and the matrix $A_k = \nabla b(\zeta_k)$ has two eigenvalues: $\lambda_k^+ > 0$ and $-\lambda_k^- < 0$.

2. The flow S generated by b admits an heteroclinic structure in $\bar{D}$. We give the technical description of this assumption. For each critical point $z_k \in \mathcal{Z}$, let $\mathcal{W}_k^s$ be the 1-dimensional locally stable manifold and $\mathcal{W}_k^u$ the 1-dimensional locally unstable manifold. Take a $\delta > 0$ small enough so that normal form conjugation and Hadamard–Perron invariant manifold theorem holds for a ball $B_k = B_\delta(\zeta_k)$ of radius $\delta > 0$ centered at each critical point . Denote $\{q_k^+, q_k^-\} = \mathcal{W}_k^u \cap B_\delta(\zeta_k)$. The hypothesis that b admits an heteroclinic structure means that for each integer $1 \le k \le N$, there is an integer $n_k^\pm \in \{1, ..., N\}$ such that

$$\lim_{t\to\infty} S^t q_k^\pm = \zeta_{n_k^\pm}.$$

3. All non degeneracy assumption made in Section 1.2 hold for each critical point.

Suppose the starting point for the diffusion $X_\epsilon(0)$ is a deterministic point in $\mathcal{W}_1^s$. Theorem 9 implies that with high probability the diffusion will exit B_1 approximately along q_1^+ or q_1^- with equal probability. Moreover, Theorem 9 tells us how to compute the scaling exponent of the additive exit correction term and the asymptotic distribution of this correction term.

For this discussion, we suppose that the diffusion exits B_1 asymptotically close to q_1^+. The exit from B_1 will now be the initial condition in Lemma 12. Applying this lemma (with $\Psi_\epsilon \equiv 0$) for sufficiently large T, we can derive the asymptotic representation of the entrance distribution for $B_{n_1^+}$, which satisfies the properties imposed to the initial condition in Theorem 9. Observe that Lemma 12 also implies that the scaling exponent of the additive correction in the entrance distribution for $B_{n_1^+}$ is the same as in the exit distribution for B_1. Moreover, this lemma implies that the asymptotic distribution of the additive correction term has in the entrance to $B_{n_1^+}$ is the evolved (under the linearization of S) version of the asymptotic distribution of the correction term at the exit from B_1. In particular, any bias on the exit of B_1 gets translated, by the linearization of the flow, to the entrance of $B_{n_1^+}$. Let $i_1 = n_1^+$. Then Theorem 9 applies again to derive that asymptotically the exit distribution from $B_{n_1^+}$ is concentrated mostly along $q_{i_1^+}$ or $q_{i_1^-}$, but with possible unequal probability. We can proceed like this iteratively along any sequence of saddle points $z_1, z_{i_1}, ..., z_{i_r}$ such that for any j, $i_{j+1} = n^+(i_j)$ or $i_{j+1} = n^-(i_j)$.

The result of this procedure allows us to conclude that the system evolves in a Markov fashion (choosing the next saddle with probability 1/2 independently of the history of the process) until it meets a saddle point at which the exit distribution becomes asymmetric. After that the choice of the two heteroclinic connections is not Markov anymore. The choice of the two heteroclinic connections may become Markov again if the system meets a saddle in which the symmetry is reestablished. We will illustrate how this phenomenon affects the exit distribution.

Planar heteroclinic network with two nodes

Let us give a concrete example. Consider the system in which b has two critical points $\{\zeta_1, \zeta_2\}$ and phase space of the flow S is as depicted in Figure 1.4, with D being a rectangle around the two critical points. Let $X_\epsilon(0) = x_0$ be on the locally stable manifold $\mathcal{W}_1^s$ of ζ_1. Consider $\{q_k^+, q_k^-\} = \mathcal{W}_k^u \cap B_\delta(\zeta_k)$,

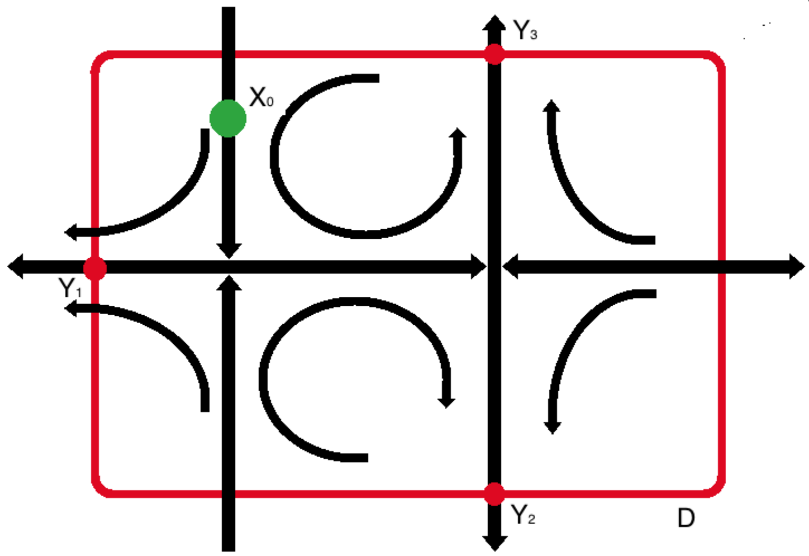

Figure 1.4: Heteroclinic network example

where $B_\delta(\zeta_k)$ is a ball of sufficiently small radius around ζ_k and $\mathcal{W}_k^u$ the locally unstable manifold of ζ_k. There is an orbit connecting ζ_1 with ζ_2:

$$\lim_{t \to \infty} S^t q_1^- = \zeta_2.$$

Recall from Section 1.1 that as $\epsilon \to$ the exit distribution concentrates near the minimizers of $V(x, \cdot)$ over the boundary of ∂D. For a heteroclinic network this means that the exit concentrates at all points in the boundary that can be reached from x along a sequence of heteroclinic connections. In this case, these points are

$$y_1 = \lim_{t \to \infty} S^t q_1^+, \quad y_2 = \lim_{t \to \infty} S^t q_2^+,$$

and $y_3 = \lim_{t\to\infty} S^t q_2^-$. Then, the exit measure will weakly converge, as $\epsilon \to 0$, to

$$p_1\delta_{y_1} + p_2\delta_{y_2} + p_3\delta_{y_3}, \tag{1.24}$$

where p_1, p_2 and p_3 are positive numbers that sum up to 1. In the spirit of Theorem 9 a direct application of our results imply a scaling limit to this convergence.

Lemma 16 *Consider the system just described. There is a family of random variables $(\theta_\epsilon)_{\epsilon>0}$, a family of random vectors $(\phi_\epsilon)_{\epsilon>0}$ and the random variable α_0 such that $\mathbf{P}\{\theta_\epsilon \in \{1,2,3\}\} = \mathbf{P}\{\alpha_0 \in (0,1]\} = 1$, and*

$$X_\epsilon(\tau_\epsilon^D) = y_{\theta_\epsilon} + \epsilon^{\alpha_0}\phi_\epsilon,$$

for every $\epsilon > 0$. The random vector $(\theta_\epsilon, \phi_\epsilon)$ converges in distribution to (θ_0, ϕ_0).

The random vector (θ_0, ϕ_0) in principle can be obtained explicitly. It is clear from (1.24) that $p_i = \mathbf{P}\{\theta_0 = i\}$, $i = 1, 2, 3$.

Let $\{\lambda_+, -\lambda_-\}$ be the set of eigenvalues of $\nabla b(\zeta_1)$ and $\{\mu_+, -\mu_-\}$ the set of eigenvalues of $\nabla b(\zeta_2)$ (see Figure 1.4). Using the three observations made at the beginning of Section 1.2.2 several cases can be considered (see [6] for further explanation):

- If $\mu_+ < \mu_-$, and $\lambda_+ < \lambda_-$ then the system is symmetric with $p_1 = 1/2$, $p_2 = p_3 = 1/4$, and $\alpha_0 = 1$. Here symmetric means that the random vector ϕ_0 has no bias along the direction of $\mathcal{W}_1^s$ if $\theta_0 = 1$ or along the direction $\mathcal{W}_1^u$ if $\theta_0 \in \{2, 3\}$. Asymmetric means that there is a bias in any of the aforementioned cases.

- If $\mu_+\lambda_+ < \mu_-\lambda_-$, and $\lambda_+ \geq \lambda_-$, the system is symmetric if $\theta_0 \in \{2, 3\}$, strongly asymmetric if $\lambda_+ < \lambda_-$ and $\theta_0 = 1$, and asymmetric if $\lambda_+ = \lambda_-$ and $\theta_0 = 1$. Moreover, $p_1 = 1/2$, $p_2 = 0, p_3 = 1/2$, when $\lambda_- < \lambda_+$, and $p_1 = 1/2, p_2 = 0, p_3 = 1/2$, when $\lambda_- = \lambda_+$. The random variable α_0 is given by
$$\alpha_0 = \frac{\lambda_-}{\lambda_+}\delta_{\{1\}}(\theta_0) + \delta_{\{2,3\}}(\theta_0).$$

- If $\mu_+\lambda_+ > \mu_-\lambda_-$, and $\lambda_+ > \lambda_-$, the system is strongly asymmetric and $p_1 = 1/2$, $p_2 = 0, p_3 = 1/2$, and
$$\alpha_0 = \frac{\lambda_-}{\lambda_+}\delta_{\{1\}}(\theta_0) + \mu_-\lambda_-/(\mu_+\lambda_+)\delta_{\{2,3\}}(\theta_0).$$

- If $\mu_+ = \mu_-$ and $\lambda_+ = \lambda_-$, the system is asymmetric and $p_1 = 1/2$, $p_2 \in (0, p_3), p_3 < 1/2$, and $\alpha_0 = 1$.

- If $\mu_+ > \mu_-$, and $\lambda_+ = \lambda_-$, the system is asymmetric if $\theta_0 = 1$ and strongly asymmetric otherwise. Moreover, $p_1 = 1/2$, $p_2 \in (0, p_3), p_3 < 1/2$, and
$$\alpha_0 = \delta_{\{1\}}(\theta_0) + (\mu_-/\mu_+)\delta_{\{2,3\}}(\theta_0).$$
- If $\mu_+ = \mu_-$, and $\lambda_+ > \lambda_-$, the system is strongly asymmetric if $\theta_0 = 1$, and asymmetric otherwise. Moreover, $p_1 = 1/2$, $p_2 = 0, p_3 = 1/2$, and $\alpha_0 = \lambda_-/\lambda_+$.

A formalization of this argument based on a weak convergence result is done in [4]. In such, the limiting behavior of the rescaled process
$$Z_\epsilon(t) = X_\epsilon(t \log(\epsilon^{-1}))$$
is obtained. Notice how this rescaled process instantaneously jumps along saddles. Hence if a weak convergence result has to be established, we need to introduce a new topology. Indeed, the standard Skorokhod topology does not allow to capture the curves along which the jumps are made. We state the weak convergence result in the next section.

Weak convergence result

In order to present the weak convergence result for the rescaled version of X_ϵ, we need to introduce a new topology.

Consider all paths $\gamma : [0,1] \to [0,\infty) \times \mathbb{R}^2$ such that the first coordinate γ^0 is nondecreasing. Equip the space of paths with the equivalence relation $\sim$, where $\gamma_1 \sim \gamma_2$ if and only if there is a path γ^* and non-decreasing surjective functions $\lambda_1, \lambda_2 : [0,1] \to [0,1]$ such that $\gamma_i = \gamma^* \circ \lambda_i$. The set of curves $\mathbf{X}$ is the quotient of the space of paths with the equivalence relation $\sim$. Actually the set $\mathbf{X}$ can be regarded as a Polish space:

Lemma 17 ([4]) $\mathbf{X}$ *can be made into a metric Polish space with distance function*
$$\rho(\Gamma_1, \Gamma_2) = \inf_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} \sup_{s \in [0,1]} |\gamma_1(s) - \gamma_2(s)|.$$

Refer to [4] for more information about this sapce.

In order to state the result, we give a non-technical introduction to the notion of entrance-exit maps introduced in [4]. Let $\mathcal{P}$ be the set of probability measures in $\mathbb{R}^2$, define $\mathrm{out} = (0,\infty) \times [0,1] \times \mathbb{R}^2 \times (0,1] \times \mathcal{P}$ and
$$\begin{aligned}\mathrm{Out_k} = \{((\mathrm{t_-}, \mathrm{p_-}, \mathrm{x_-}, \beta_-, \mathrm{F_-}), (\mathrm{t_+}, \mathrm{p_+}, \mathrm{x_+}, \beta_+, \mathrm{F_+})) \in \mathrm{out}^2 :& \\ t_- = t_+, x_\pm = q_k^\pm, p_- + p_+ = 1, \beta_- = \beta_+\}&.\end{aligned}$$

Then we have the following definitions.

Definition 18 *For each k, an entrance-exit map is a map*

$$\Psi_k : \{q_k^+, q_k^-\} \times (0, 1] \times \mathcal{P} \to \mathrm{Out_k},$$

where the domain of Ψ_k satisfies some regularity assumptions(see [4, page 10]) . We denote $\Psi_k = (\Psi_k^+, \Psi_k^-)$.

Definition 19 *Suppose $x_0 \in \mathcal{W}_k^s$, for some $1 \leq k \leq N$. The sequence $\mathbf{z} = (\theta_0, z_{i_1}, ..., \theta_{r-1}, z_{i_r}, \theta_r)$ is admissible for x_0 (refered as x_0-admissible) if*

1. *θ_0 is the orbit of x_0 with $S^t x_0 \to z_{i_1}$, as $t \to \infty$;*
2. *for each $j \in \{1, .., r\}$, θ_j is either the orbit of $q_{i_j}^+$ or the orbit $q_{i_j}^-$;*
3. *for each $j \in \{1, .., r\}$, i_{j+1} is either $n_{i_j}^+$ or $n_{i_j}^-$ according to whether theta$_j$ is the orbit of $q_{i_j}^+$ or $q_{i_j}^-$.*

With each admissible sequence $\mathbf{z}$ we associate the sequence

$$\eta(\mathbf{z}) = ((x_0', \alpha_0, \mu_0), (t_1, p_1, x_1, \alpha_1, \mu_1), ..., (t_r, p_r, x_r, \alpha_r, \mu_r)),$$

where $x_0' = S^{t'(x_0)} x_0, \alpha_0 = 1$,

$$t'(x_0) = \inf\{t > 0 : S^t x_0 \in B_\delta(\zeta_{i_1})\},$$

and the rest of the entries are given by

$$(t_j, p_j, x_j, \alpha_j, \mu_j) = \begin{cases} \Psi_{i_j}^+(x_{j-1}, \alpha_{j-1}, \mu_{j-1}), & i_j = n_{i_j}^+ \\ \Psi_{i_j}^-(x_{j-1}, \alpha_{j-1}, \mu_{j-1}), & i_j = n_{i_j}^- \end{cases} .$$

To each admissible sequence $\mathbf{z}$ we can associate a piecewise constant curve $\Gamma(\mathbf{z})$ by identifying it with the path of curves such that spend time t_j at the point x_j and jump to the next point along the path θ_j. Also we can associate probabilities through the relationship

$$\pi(\mathbf{z}) = p_1 ... p_r.$$

Note how the set of all admissible sequences for x_0 has the structure of a binary tree partially ordered by inclusion. We say that a set of admissible sequences L of x_0 is free if no two sequences of L are comparable with respect to this partial order. Additionally, if any sequence not in L is comparable to one sequence from L then L is called complete. It is clear that for any free set $\pi(L) := \sum_{\mathbf{z} \in L} \pi(\mathbf{z}) \leq 1$, while for a complete set $\pi(L) = 1$.

The main theorem is:

Theorem 20 *Suppose that $X_\epsilon(0) = x_0$ is in the heteroclinic invariant. For each $\epsilon > 0$, define the process $Z_\epsilon(t) = X_\epsilon(t|\log(\epsilon)|)$. Then, for any conservative set L of x_0-admissible sequences, there is a family of stopping times $(T_\epsilon)_{\epsilon>0}$ such that the distribution of the graph $\Gamma_{Z_\epsilon(t),t<T_\epsilon}$ converges weakly in $(\mathbb{X}, \rho)$ to the measure $M_{x_0,L}$ concentrated on the set*

$$\{\Gamma(\mathbf{z}) : \mathbf{z} \in L\}$$

and satisfying $M_{X_\epsilon(0),L}\{\Gamma(\mathbf{z})\} = \pi(\mathbf{z})$.

Contributions made in the case S admits a heteroclinic network

In this section we outline our contribution for the case in which S admits an heteroclinic network.

The iteration procedure described in Section 1.4.2 was first proposed in [4]. It is the central idea in proving the main results in [4]. This iteration is carried out in [4] by using an equivalent version of Theorem 9. This version was proved under the hypothesis that the non-linear system can be locally conjugated to a linear system by a C^2 transformation. That is, in [4] Theorem 20 is proved under the following hypothesis:

Condition 21 *At each critical point $\zeta \in \mathcal{Z}$ there are non-resonant conditions.*

As discussed in Section 1.2 this is in general not the case, and examples of saddle point that do not satisfy this condition are known [48]. In this work, we completely remove condition 21 in the 2-dimensional situation.

On the other hand, observe that the iteration procedure is Section 1.4.2 is based on the computation of a random map. This map is such that, for a domain V, to any given initial distribution of the diffusion X_ϵ, it gives the exit distribution of X_ϵ from V. We call this map a random Poincaré map. For $V \subset \mathbb{R}^2$, let Π_V be the set of probability measures with support on V. Then, the random Poincaré map for D, $\Upsilon_D : \Pi_D \to \Pi_{\partial D}$ is the (deterministic) map such that $\Upsilon_D\mathbf{Q} = \mathbf{P_Q}\{X_\epsilon(\tau_\epsilon^D\}$, where $\mathbf{P_Q}$ is the original probability measure conditioned on $X_\epsilon(0)$ being distributed as $\mathbf{Q}$. The iteration in Section 1.4.2 illustrates the use of this map. Notice that this methodology applies regardless the type of equilibria that the system exhibits. We base our proof of Theorems 9 and 13 on a similar idea. Hence, it is worth to study small noise perturbations with this direction in mind.

As an example of a Poincaré map, consider our example in Section 1.4.2. The exit distribution is the composition of the Poincaré maps $\Upsilon_{D_5} \circ ... \circ \Upsilon_{D_1}\delta_{x_0}$, where D_i are ilustrated in Figures 1.5 and **??**, and we are conditioning on exit along y_3.

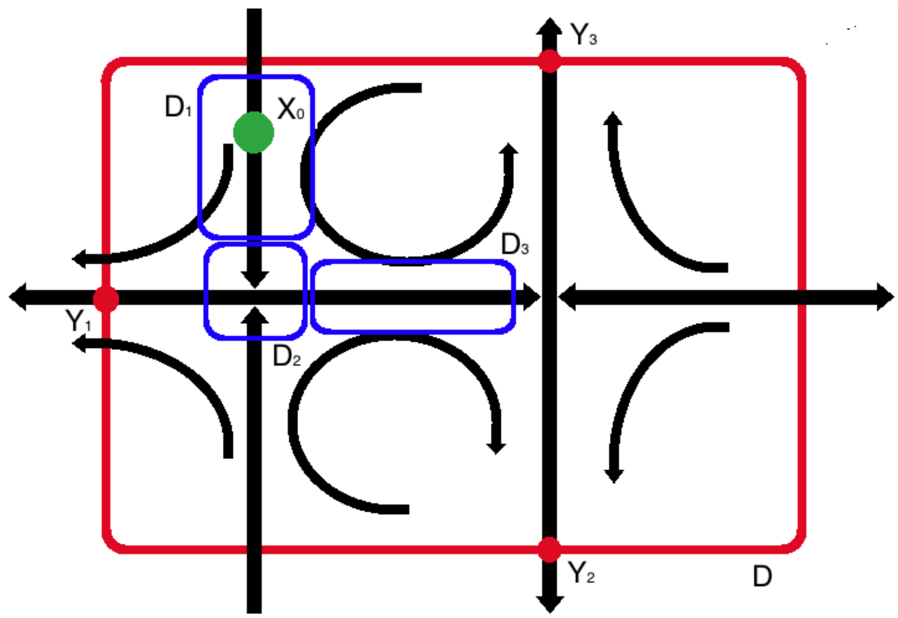

Figure 1.5: Illustration of the domains D_1, D_2, and D_3 used to compute the Poincaré maps in the case of a heteroclinic network with 2 nodes conditioned on exit along y_3: escape from the first saddle.

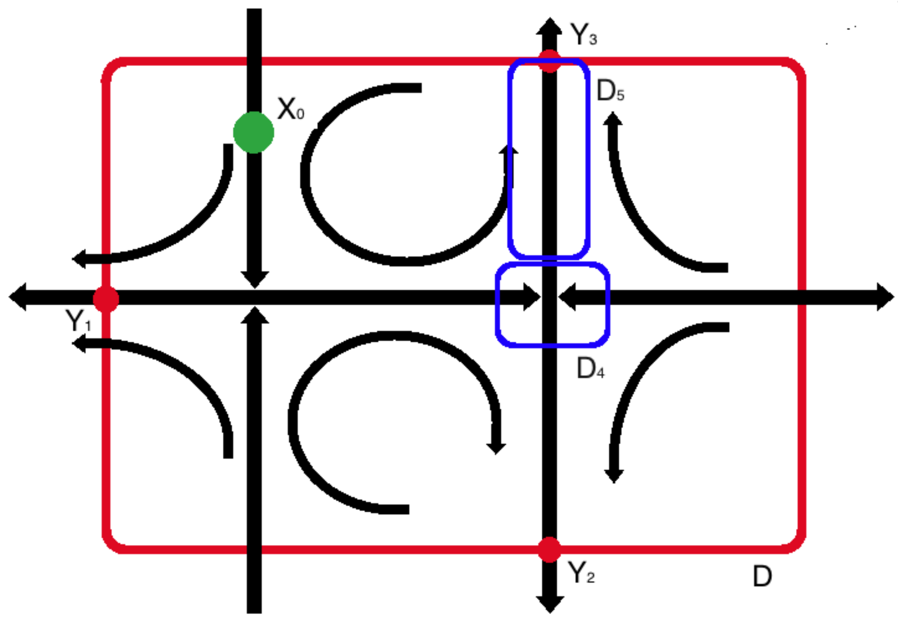

Figure 1.6: Illustration of the domains D_3 and D_4 used to compute the Poincaré maps in the case of a heteroclinic network with 2 nodes condition on exit along y_3: escape from the second saddle.

1.5 General Setting

The objective of this section is to establish the general setting and notation, although each chapter has the necessary modifications and additions to the following.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a complete probability space (every subset of every measurable null set is measurable) and W be a d-dimensional standard Brownian Motion on it. Let $(\mathcal{F}_t^W)_{t>0}$ be the filtration generated by W which satisfies the usual hypothesis [54, Section I.5] . We assume that $(\Omega, \mathcal{F}, \mathbf{P})$ is rich enough to accommodate a family of random vectors $(\xi_\epsilon)_{\epsilon \geq 0}$ in $\mathbb{R}^d$ such that the sigma algebra generated by ξ_ϵ is independent of $\mathcal{F}_\infty^W$ for each $\epsilon > 0$. For each $\epsilon > 0$, we consider the left continuous filtration

$$\mathcal{G}_t^\epsilon = \sigma(\xi_\epsilon) \vee \mathcal{F}_t^W,$$

as well as the collection of null sets

$$N_0 = \{Z \subset \Omega : \exists G \in \mathcal{G}_\infty^\epsilon \text{ with } Z \subset G \text{ and } \mathbf{P}\{G\} = 0\}.$$

Let us create the augmented filtration $\mathcal{F}_t^\epsilon = \sigma(\mathcal{G}_t^\epsilon \cup N_0)$ for $t \in [0, \infty)$, and $\mathcal{F}_\infty^\epsilon = \sigma(\cup_{t\geq 0} \mathcal{F}_t^\epsilon)$. It can be shown that W is a brownian motion with respect to $(\mathcal{F}_t^\epsilon)_{t\geq 0}$, the path of W is independent of ξ_ϵ and $(\mathcal{F}_t^\epsilon)_{t\geq 0}$ satisfies the usual hypothesis, for every $\epsilon > 0$.

Throughout the text, we suppose that the family of random variables $(\xi_\epsilon)_{\epsilon\geq 0}$ satisfies $\xi_\epsilon \to \xi_0$ in distribution, and that ξ_ϵ has a finite second moment for each $\epsilon > 0$.

Consider a C^∞-smooth vector field b on $\mathbb{R}^d$ and a C^2-smooth matrix valued function $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$. Consider the Itô stochastic differential equation

$$dX_\epsilon = b(X_\epsilon)dt + \epsilon\sigma(X_\epsilon)dW \tag{1.25}$$

equipped with initial condition

$$X_\epsilon(0) = x_0 + \epsilon^\alpha \xi_\epsilon, \tag{1.26}$$

where $\alpha \in (0, 1]$. Hypothesis regarding the point $x_0 \in \mathbb{R}^d$ will be given in each chapter. We assume that both b and σ are uniformly Lipschitz and bounded, i.e., there is a constant $L > 0$ such that

$$|\sigma(x) - \sigma(y)| \vee |b(x) - b(y)| \leq L|x - y|, \quad x, y \in \mathbb{R}^2,$$
$$|\sigma(x)| \vee |b(x)| \leq L, \quad x \in \mathbb{R}^2,$$

where $|\cdot|$ denotes the Euclidean norm for vectors and Hilbert–Schmidt norm for matrices. Further assume that the matrix function $a = \sigma\sigma^*$ is uniformly positive definite. These conditions can be weakened, but we prefer this setting to avoid multiple localization procedures throughout the text. These assumptions imply [41, Theorems 5.2.5 and 5.2.9] that equation (1.25) has a strong solution with strong uniqueness on the filtered probability space $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}^\epsilon_t)_{t\geq 0})$ with initial condition (1.26) for each $\epsilon > 0$. Let us recall the definition of strong uniqueness and strong solution for completeness.

Definition 22 *A strong solution of the stochastic differential equation* (1.25) *with initial condition* (1.26) *on the filtered probability space* $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}^\epsilon_t)_{t\geq 0})$ *is a process* $X_\epsilon = \{X_\epsilon(t); 0 \leq t < \infty\}$ *with continuous sample paths and with the following properties:*

1. X_ϵ *is adapted to the filtration* $(\mathcal{F}^\epsilon_t)_{t\geq 0}$,
2. $\mathbf{P}\{X_\epsilon(0) = x_0 + \epsilon^\alpha \xi_\epsilon\} = 1$,
3. $\mathbf{P}\{\int_0^t (|b_i(X_\epsilon(s))| + \sigma_{i,j}(X_\epsilon(s))^2) ds < \infty\} = 1$ *for every* $1 \leq i, j \leq d$ *and* $t \geq 0$,
4. *the integral version of* (1.25)
$$X_\epsilon(t) = X_\epsilon(0) + \int_0^t b(X_\epsilon(s))ds + \epsilon \int_0^t \sigma(X_\epsilon(s))dW(s); 0 \leq t < \infty,$$
holds with probability 1.

Given two strong solutions X_ϵ *and* $\widetilde{X}_\epsilon$ *of* (1.25) *with initial condition* (1.26) *relative to the same brownian motion* W. *Then, we say that strong uniqueness holds whenever* $\mathbf{P}\{X_\epsilon(t) = \widetilde{X}_\epsilon(t); 0 \leq t < \infty\} = 1$.

For a general background on stochastic differential equations see, for example, [41, Chapter 5].

The flow generated by b is denoted by $S = (S^t x)_{(t,x)\in\mathbb{R}\times\mathbb{R}^d}$. That is, $S^t x$ satisfies
$$\frac{d}{dt} S^t x = b(S^t x), \quad S^0 x = x.$$
The linearization of S along the orbit of x is denoted by $\Phi_x(t)$:
$$\frac{d}{dt}\Phi_x(t) = A(t)\Phi_x(t), \;\; \Phi_x(0) = I, \tag{1.27}$$

where $A(t) = \nabla b(S^t x)$ and I is the identity matrix. Here ∇ is the derivative operator, that is, for a differentiable vector field $h : \mathbb{R}^d \to \mathbb{R}^d$, ∇h is the $\mathbb{R}^{d\times d}$ matrix derivative of h.

Throughout the text D is a domain (open, connected and bounded) in $\mathbb{R}^d$ with piecewise C^2 boundary.

The exit problem for the diffusion process X_ϵ in D is studied. We are interested in the joint asymptotic properties (as $\epsilon \to 0$) of $(X_\epsilon(\tau_\epsilon^D), \tau_\epsilon^D)$, where τ_ϵ^D is the stopping time defined by

$$\tau_\epsilon^D = \tau_\epsilon^D(x_0) = \inf\{t > 0 : X_\epsilon(t) \in \partial D\}.$$

.

Specific hypotheses on the vector field will be given in each chapter. On the other hand, the abstract formulation will not be given in each chapter, instead we assume this technical formulation to hold throughout the text.

1.5.1 Organization of the Text

The organization of the rest of the text closely mimics the presentation given in this chapter.

In Chapter 2 the planar (i.e. $d = 2$) exit problem is studied under the assumption that S has a unique saddle at the origin. That is, $0 \in \mathbb{R}^2$ is the only critical point $b(0) = 0$ and the eigenvalues λ_+, $-\lambda_-$ of the of the matrix $\nabla b(0)$ are such that $\lambda_\pm > 0$. The exit problem is studied conditioned that the process X_ϵ starts on the stable manifold $\mathcal{W}^s$ of 0.

In Chapter 3 the Levinson case in arbitrary dimensions is considered. We also proved Lemma 12 stated in this chapter, and use it intesively in the proof of the Levinson case. In this section, we also study the 1-dimensional example discussed in Section 1.4.1 of this chapter.

In Chapter 4 we present a short survey on how the techniques in this text can be extended. Several open problems are also discussed.

Chapter 2

Saddle Point

In this chapter we study the stochastic process X_ϵ when the underlying deterministic system S has a unique saddle point.

In Section 2.1 we introduce the setting, which relies on the setting presented in Section 1.5 of Chapter 1. In Section 2.2 we state the main theorem and split the proof into several parts. In Section 2.3 we introduce a simplifying change of coordinates in a small neigborhood of the saddle point. The analysis of the transformed process in Section 2.4 is based upon two results. Their proofs are given in Sections 2.5 and 2.6.

2.1 Setting

For this chapter we consider the general formulation made in Section 1.5 of Chapter 1, except that we restrict ourselves to the 2−dimensional situation. The process X_ϵ is the strong solution of (1.25), under the assumptions made on the C^∞-smooth vector field b, the C^2-smooth matrix valued function $\sigma : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ and the standard 2-dimensional Wiener process W in Section 1.5 of Chapter 1.

We will study the exit problem from the domain $D \subset \mathbb{R}^2$ with piecewise C^2 boundary. Assume that the origin 0 belongs to D and it is a unique fixed point for S in $\bar{D}$, or, equivalently, a unique critical point for b in $\bar{D}$. Therefore,

$$b(x) = Ax + Q(x),$$

where $A = \nabla b(0)$ and Q is the non-linear part of the vector field satisfying $|Q(x)| = O(|x|^2)$, $x \to 0$.

Suppose that 0 is a hyperbolic critical point, i.e. the matrix A has two eigenvalues λ_+ and $-\lambda_-$ satisfying $-\lambda_- < 0 < \lambda_+$. Without loss of gen-

erality, we suppose that the canonical vectors are the eigenvectors for the matrix, so that $A = \mathrm{diag}(\lambda_+, -\lambda_-)$.

For an interval $J \subset \mathbb{R}$, let $S^J x$ denote the set

$$S^J x = \{S^t x : t \in J\}.$$

According to the Hadamard–Perron Theorem (see e.g. [53, Section 2.7]), the curves $\mathcal{W}^s$ and $\mathcal{W}^u$ defined via

$$\mathcal{W}^u = \{x \in \bar{D} : \lim_{t\to-\infty} S^t x = 0, \text{ and for some } s \geq 0, S^{(-\infty,s)}x \subset D \text{ and } S^{(s,\infty)}x \cap \bar{D} = \emptyset\},$$

and,

$$\mathcal{W}^s = \{x \in \bar{D} : \lim_{t\to\infty} S^t x = 0, \text{ and for some } s \leq 0, S^{(-\infty,s)}x \cap \bar{D} = \emptyset \text{ and } S^{(s,\infty)}x \subset D\}.$$

are smooth, invariant under S and tangent to e_2 and, respectively, to e_1 at 0. The curve $\mathcal{W}^s$ is called the stable manifold of 0, and $\mathcal{W}^u$ is called the unstable manifold of 0.

We assume that $\mathcal{W}^u$ intersects ∂D transversally at points q_+ and q_- such that the segment of $\mathcal{W}^u$ connecting q_- and q_+ lies entirely inside D and contains 0.

We fix a point $x_0 \in \mathcal{W}^s \cap D$ and equip (1.25) with the initial condition

$$X_\epsilon(0) = x_0 + \epsilon^\alpha \xi_\epsilon, \quad \epsilon > 0, \tag{2.1}$$

where $\alpha \in (0, 1]$ is fixed, and $(\xi_\epsilon)_{\epsilon>0}$ is a family of random vectors independent of W, such that for some random vector ξ_0, $\xi_\epsilon \to \xi_0$ as $\epsilon \to 0$ in distribution.

If $\alpha \neq 1$, then we impose a further technical condition

$$\mathbf{P}\{\xi_0 \parallel b(x_0)\} = 0, \tag{2.2}$$

where $\parallel$ denotes collinearity of two vectors.

2.2 Main Result.

The main result of the present chapter is the following:

Theorem 23 *In the setting described above, there is a family of random vectors $(\phi_\epsilon)_{\epsilon>0}$, a family of random variables $(\psi_\epsilon)_{\epsilon>0}$, and a number*

$$\beta = \begin{cases} 1, & \alpha\lambda_- \geq \lambda_+ \\ \alpha\frac{\lambda_-}{\lambda_+}, & \alpha\lambda_- < \lambda_+ \end{cases} \tag{2.3}$$

such that

$$X_\epsilon(\tau_\epsilon^D) = q_{\mathrm{sgn}(\psi_\epsilon)} + \epsilon^\beta \phi_\epsilon.$$

The random vector

$$\Theta_\epsilon = \left(\psi_\epsilon, \phi_\epsilon, \tau_\epsilon^D + \frac{\alpha}{\lambda_+}\ln\epsilon\right)$$

converges in distribution as $\epsilon \to 0$.

The distribution of ψ_ϵ,ϕ_ϵ, and the distributional limit of Θ_ϵ will be described precisely.

The proof of Theorem 23 has essentially three parts involving the analysis of diffusion (i) along $\mathcal{W}^s$; (ii) in a small neighborhood of the origin; (iii) along $\mathcal{W}^u$.

In order to study the first part, we need to introduce $\Phi_x(t)$ as the linearization of S along the orbit of $x \in \mathbb{R}^2$, i.e. we define $\Phi_x(t)$ to be the solution to the matrix ODE

$$\frac{d}{dt}\Phi_x(t) = A(t)\Phi_x(t), \ \ \Phi_x(0) = I,$$

where $A(t) = \nabla b(S^t x)$. We have the following theorem:

Theorem 24 *Let* $x \in \mathbb{R}^2$ *and* $(\xi_\epsilon)_{\epsilon>0}$ *be a family of random vectors independent of* W *and convergent in distribution, as* $\epsilon \to 0$, *to* ξ_0. *Suppose* $\alpha \in (0,1]$ *and let* X_ϵ *be the solution of the SDE* (1.25) *with initial condition* $X_\epsilon(0) = x + \epsilon^\alpha \xi_\epsilon$. *Then, for every* $T > 0$, *the following representation holds true:*

$$X_\epsilon(T) = S^T x + \epsilon^\alpha \bar{\xi}_\epsilon, \quad \epsilon > 0,$$

where

$$\bar{\xi}_\epsilon \xrightarrow{Law} \bar{\xi}_0, \quad \epsilon \to 0,$$

with

$$\bar{\xi}_0 = \Phi_x(T)\xi_0 + \mathbf{1}_{\{\alpha=1\}} N,$$

N being a Gaussian vector:

$$N = \Phi_x(T)\int_0^T \Phi_x(s)^{-1}\sigma(S^s x)dW(s).$$

If $\alpha = 1$ *or assumption* (2.2) *holds, then* $\mathbf{P}\{\bar{\xi}_0 \parallel b(S^T x)\} = 0$.

The second part of the analysis is the core of the chapter. Theorem 25 below describes the behavior of the process in a small neighborhood U of the origin. Notice that since $x_0 \in \mathcal{W}^s$, one can choose T large enough to ensure that that $S^T x_0 \in \mathcal{W}^s \cap U$. Therefore, the conditions of the following result are met if we use the terminal distribution of Theorem 24 (applied to the initial data given by (2.1)) as the initial distribution.

Theorem 25 *There are two neighborhoods of the origin $U \subset U' \subset D$, two positive numbers $\delta < \delta'$, and C^2 diffeomorphism $f : U' \to (-\delta', \delta')^2$, such that $f(U) = (-\delta, \delta)^2$ and the following property holds:*

Suppose $x \in \mathcal{W}^s \cap U$, and $(\xi_\epsilon)_{\epsilon>0}$ is a family of random variables independent of W and convergent in distribution, as $\epsilon \to 0$, to ξ_0, where ξ_0 satisfies (2.2) with respect to x. Assume that $\alpha \in (0, 1]$ and that X_ϵ solves (1.25) with initial condition

$$X_\epsilon(0) = x + \epsilon^\alpha \xi_\epsilon, \tag{2.4}$$

where ξ_ϵ satisfies condition (2.2) with respect to x.

There is also a family of random vectors $(\phi'_\epsilon)_{\epsilon>0}$, and a family of random variables $(\psi'_\epsilon)_{\epsilon>0}$, such that

$$X_\epsilon(\tau^U_\epsilon) = g(\operatorname{sgn}(\psi'_\epsilon)\delta e_1) + \epsilon^\beta \phi'_\epsilon,$$

where $g = f^{-1}$, β is defined in (2.3), and the random vector

$$\Theta'_\epsilon = \left(\psi'_\epsilon, \phi'_\epsilon, \tau^U_\epsilon + \frac{\alpha}{\lambda_+} \ln \epsilon\right)$$

converges in distribution as $\epsilon \to 0$.

The notation for Θ'_ϵ and its components is chosen to match the notation involved in the statement of Theorem 23. Random elements ψ'_ϵ,ϕ'_ϵ and the distributional limit of Θ'_ϵ will be described precisely, see (2.30). Obviously, the symmetry or asymmetry in the limiting distribution of ψ'_ϵ results in the symmetric or asymmetric choice of exit direction so that the exits in the positive and negative directions are equiprobable or not. On the other hand, the limiting distribution of ϕ'_ϵ determining the asymptotics of the exit point can also be symmetric or asymmetric which results in the corresponding features of the random choice of the exit direction at the next saddle point visited by the diffusion.

In Section 2.4 we prove Theorem 25 using the approach based on normal forms.

The last part of the analysis is devoted to the exit from D along $\mathcal{W}^u$. We need the following statement which is a specific case of the main result of Chapter 3.

Theorem 26 *In the setting of Theorem 24, assume additionally that (i) $q = S^T x \in \partial D$; (ii) there is no $t \in [0, T)$ with $S^t x \in \partial D$; (iii) $b(q)$ is tranversal (i.e. not tangent) to ∂D at q. Then*

$$\tau_\epsilon^D \xrightarrow{\mathbf{P}} T, \quad \epsilon \to 0, \tag{2.5}$$

and

$$\epsilon^{-\alpha}(X_\epsilon(\tau_\epsilon^D) - q) \xrightarrow{Law} \pi\bar{\xi}_0, \quad \epsilon \to 0, \tag{2.6}$$

where π denotes the projection along $b(q)$ onto the tangent line to ∂D at q.

Now Theorem 23 follows from the consecutive application of Theorems 24 through 26 and with the help of the strong Markov property. In fact, in this chain of theorems, the conclusion of Theorem 24 ensures that the conditions of Theorem 25 hold, and the conclusion of the latter ensures that the conditions of Theorem 26 hold. Notice that the total time needed to exit D equals the sum of times described in the three theorems. Notice also that at each step we can compute the limiting initial and terminal distributions explicitly. Theorems 24 and 26 contain the respective formulas in their formulations, and the explicit limiting distribution for Θ'_ϵ of Theorem 25 is computed in (2.30).

2.3 Simplifying change of coordinates

2.3.1 Smooth Transformation and Normal Forms

In this section we give a brief review of the theory of Normal Forms. In particular, we focus on the neighborhood of a saddle point for the deterministic flow S.

The idea is to find a local change of variables $\theta : \mathbb{R}^2 \to \mathbb{R}^2$ such that $z(t) = \theta(S^t x)$ satisfies $\dot{z} = Az$ with the appropriate initial condition. First, note that z satisfies the equation

$$\begin{aligned} \frac{d}{dt} z(t) &= \nabla\theta(S^t x) b(S^t x) \\ &= \nabla\theta(z(t))^{-1} b(\theta^{-1}(z(t))), \quad z(0) = \theta(x). \end{aligned}$$

Hence, the goal is to find a transformation $\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that leaves $\nabla\theta(z)^{-1}b(\theta^{-1}(z))$ as simple as possible (ideally equal to Az).

We start with some notions. For a multi-index $\alpha = (\alpha_1, \alpha_2) \in \mathbb{Z}^2_+$ and a base $\{e_1, e_2\}$ of $\mathbb{R}^2$(as a vector space over $\mathbb{R}$) we denote the monomial $x^\alpha e_i = x_1^{\alpha_1} x_2^{\alpha_2} e_i$.

Definition 27 *For a non-negative integer r, the space of linear combinations (over $\mathbb{R}$) of monomials $x^\alpha e_i$ with $|\alpha| = \alpha_1 + \alpha_2 = r$, is called the space of Homogenous Polynomials in 2 variables of degree r. This space is denoted as $\mathcal{H}_r$. In other words, $\mathcal{H}_r$ is,*

$$\mathcal{H}_r = \mathrm{span}_{\mathbb{R}} \left\{ \mathrm{x}^\alpha \mathrm{e_j} : \alpha \in \mathbb{Z}^2_+, |\alpha| = \mathrm{r} \ \textit{and} \ 1 \leq \mathrm{j} \leq 2 \right\}.$$

It is easy to see that $\mathcal{H}_r$ is isomorphic (as a vector space over the real numbers) to $\mathbb{R}^{2(r+1)}$.

Using this notation, use Taylor's classical theorem to decompose the function $b : \mathbb{R}^2 \to \mathbb{R}^2$ as

$$b(z) = Az + b_2(z) + ... + b_R(z), \tag{2.7}$$

with $b_i \in \mathcal{H}_i$ for $1 \leq i \leq R$, and $b_R(x) = O(|x|^R)$ as $|x| \to 0$.

Suppose that $z = \theta_k(\zeta)$, where θ_k is the near identity transformation

$$\theta_k(\zeta) = \zeta + h_k(\zeta), \quad h_k \in \mathcal{H}_k, \quad k \geq 2. \tag{2.8}$$

Note that θ_k is a topological diffeomorphism in a small open neighborhood of the origin Ω_k. Throughout we restrict the analysis inside Ω_k. A Taylor approximation shows that the inverse of θ_k satisfies

$$\begin{aligned}\theta_k^{-1}(\zeta) &= \zeta - h_k(\zeta) + O(|\zeta|^{2k-1}) \\ &= \zeta - h_k(\zeta) + O(|\zeta|^{k+1}).\end{aligned} \tag{2.9}$$

Further application of Taylor's approximation together with the condition that $k \geq 2$, imply that for any $\zeta \in \Omega_k$,

$$\begin{aligned}\nabla\theta_k(\zeta)^{-1} &= I - \nabla h_k(\zeta) + O(|\zeta|^{2(k-1)}) \\ &= I - \nabla h_k(\zeta) + O(|\zeta|^k).\end{aligned}$$

Also, from (2.9), we obtain that for any $i = 1, ..., R-1$,

$$b_i(\theta_k^{-1}(\zeta)) = b_i(\zeta) + O(|\zeta|^{k+1}).$$

Using (2.7) and this bounds, we get that

$$\begin{aligned}(\nabla\theta_k(\zeta))^{-1}b(\theta_k^{-1}(\zeta)) &= A\zeta + b_2(\zeta) + ... + b_{k-1}(\zeta)\\ &\quad +(b_k(\zeta) - \mathcal{L}_A^k h_k(\zeta)) + O(|\zeta|^{k+1}),\end{aligned}$$

where we defined the operator $\mathcal{L}_A^k : \mathcal{H}_k \to \mathcal{H}_k$ by

$$\mathcal{L}_A^k h(\zeta) = h(\zeta)A\zeta - A\nabla h(\zeta). \tag{2.10}$$

It is clear that the following theorem holds:

Theorem 28 *Let $\mathcal{R}(\mathcal{L}_A^k) \subset \mathcal{H}_k$ be the range of the operator $\mathcal{L}_A^k : \mathcal{H}_k \to \mathcal{H}_k$. Take $\mathcal{I}_k \subset \mathbb{R}^2$ be any subspace such that $\mathcal{H}_k = \mathcal{R}(\mathcal{L}_A^k) \oplus \mathcal{I}_k$. Then, there is a sequence of near identity transformations of the form (2.8) and nested neighborhoods of the origin $\Omega_{k+1} \subset \Omega_k$, such that $z(t) = \theta_r \circ \cdots \circ \theta_2(S^t x)$ satisfies*

$$\frac{d}{dt}z(t) = Az(t) + b_2(z(t)) + \cdots + b_r(z(t)) + O(|z|^{r+1}),$$

inside Ω_r, and $b_k \in \mathcal{I}_k$, $k = 1, ..., r$.

An equation written in this form is said to be in Normal Form up to order r.

The idea is to characterize the image of the operator $\mathcal{L}_A^k$ and simplify each non-linear part of b, starting from b_2 and all the way up to b_R. In order to achieve this, we remark that, $x^\alpha e_j$ is an eigenvector of $\mathcal{L}_A^k$ for any $\alpha \in \mathbb{Z}_+^2$:

$$\mathcal{L}_A^k x^\alpha e_j = (\lambda^T\alpha - \lambda_j)x^\alpha e_j,$$

for $\lambda = (\lambda_+, \lambda_-)$. This motivates the following definition:

Definition 29 *A pair of complex numbers $\lambda = (\lambda_1, \lambda_2)$ is said to be non-resonant if there are no integral relations between them of the form $\lambda_j = \alpha\cdot\lambda$, where $\alpha = (\alpha_1, \alpha_2) \in \mathbb{Z}_+^2$ is a multi-index with $|\alpha| = \alpha_1 + \alpha_2 \geq 2$. Otherwise, we say that $\lambda = (\lambda_1, \lambda_2)$ is resonant.*

A resonant λ is said to be one-resonant if all the resonance relations for λ follow from a single resonance relation.

A monomial $x^\alpha e_j = x_1^{\alpha_1}x_2^{\alpha_2}e_j$ is called a resonant monomial of order R if $\alpha\cdot\lambda = \lambda_j$ and $|\alpha| = R$.

In the spirit of Theorem 28 it is clear (see [38],[20]) that for any pair of integers $R \geq 1$ and $k \geq 1$, there are two neighborhoods of the origin Ω_f and

Ω_g and a C^k-diffeomorphism $f : \Omega_f \to \Omega_g$ with inverse $g : \Omega_g \to \Omega_f$ such that

$$(\nabla g(y))^{-1}\, b(g(y)) = Ay + P(y) + \mathcal{R}(y), \quad y \in \Omega_g \tag{2.11}$$

where P is a polynomial containing only resonant monomials of order at most R and $\mathcal{R}(\zeta) = O(|\zeta|^{R+1})$. Moreover, the so called Poincaré theorem [20, Theorem 2.2.4] asserts that if λ is non-resonant, then f can be chosen so that both P and $\mathcal{R}$ in (2.11) are identically zero. If λ is one-resonant then [38, Theorem 3,Section 2] says that f can be chosen so that $\mathcal{R}$ in (2.11) is identically zero. More precisely:

Lemma 30 *For any $k \geq 1$, there are two neighborhoods of the origin Ω_f and Ω_g and a C^k-diffeomorphism $f : \Omega_f \to \Omega_g$ with inverse $g : \Omega_g \to \Omega_f$ such that*

$$(\nabla g(y))^{-1}\, b(g(y)) = Ay + P(y), \quad y \in \Omega_g, \tag{2.12}$$

where P is a polynomial that contains only resonant monomials.

This is the core result we use to study the stochastic case in the next section.

2.3.2 Change of Variables in the Stochastic Case

In this section we start analyzing the diffusion in the neighborhood of the saddle point. The first step is to find a smooth coordinate change that would simplify the system. This can be done with the help of the theory of normal forms presented on the last section.

Let g be a $C^\infty-$diffeomorphism of a neighborhood of the origin with inverse f. When X_ϵ is close to the origin and belongs to the image of that neighborhood under g, we can use Itô's formula to see that $Y_\epsilon = f(X_\epsilon)$ satisfies

$$\begin{aligned} dY_\epsilon &= \nabla f(X_\epsilon) dX_\epsilon + \frac{1}{2}[\nabla f(X_\epsilon), X_\epsilon] \\ &= \nabla f(g(Y_\epsilon)) b(g(Y_\epsilon)) dt + \epsilon \tilde{\sigma}(Y_\epsilon) dW + \epsilon^2 \Psi(Y_\epsilon) dt, \end{aligned}$$

for some smooth function $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ and $\tilde{\sigma} = ((\nabla f) \circ g)\, \sigma$. Here the square brackets mean quadratic covariation. Since $\nabla f \circ g = (\nabla g)^{-1}$, we can rewrite the above SDE as in the deterministic case as

$$dY_\epsilon = \Big((\nabla g(Y_\epsilon))^{-1}\, b(g(Y_\epsilon)) + \epsilon^2 \Psi(Y_\epsilon) \Big)\, dt + \epsilon \tilde{\sigma}(Y_\epsilon) dW. \tag{2.13}$$

In order to simplify the drift term in this equation, we rely on Lemma 30. First, note that $(\lambda_+, -\lambda_-)$ is either non-resonant or one-resonant (resonant

cases that are not one-resonant are possible in higher dimensions where pairs of eigenvalues get replaced by vectors of eigenvalues). The non-resonant case (in any dimension) was studied in [4]. In this paper, we extend the analysis of [4] to the non-resonant case, i.e. the one-resonant case, given that we are working in 2 dimensions.

To find all resonant monomials of a given order $r \geq 2$, we have to find all the integer solutions to the two 2×2 systems of equations:

$$\alpha_1\lambda_+ - \alpha_2\lambda_- = \pm\lambda_\pm,$$
$$\alpha_1 + \alpha_2 = r.$$

Therefore, the power multi-indices of a resonant monomial of order r has to coincide with one of the following:

$$(\alpha_1^+(r), \alpha_2^+(r)) = \frac{1}{\lambda_+ + \lambda_-}(\lambda_+ + r\lambda_-, (r-1)\lambda_+), \tag{2.14}$$
$$(\alpha_1^-(r), \alpha_2^-(r)) = \frac{1}{\lambda_+ + \lambda_-}((r-1)\lambda_-, r\lambda_+ + \lambda_-), \tag{2.15}$$

Let us make some elementary observations on integer solutions of these equations for $r \geq 2$.

1. None of the solution indices can be 0. Moreover, neither $\alpha_1^+(r)$ nor $\alpha_2^-(r)$ can be equal to 1.

2. As functions of r, $\alpha_i^\pm(r)$ are increasing.

3. Expressions (2.14) and (2.15) cannot be an integer for $r = 2$.

4. The term $P = (P_1, P_2)$ in (2.12) satisfies $P_1(y) = O(y_1^2|y_2|)$ and $P_2(y) = O(|y_1|y_2^2)$. This observation is a consequence of observations 1 and 3 since they imply that resonant multi-indices have to satisfy $\alpha^+(r) \geq (2,1)$ and $\alpha^-(r) \geq (1,2)$ coordinatewise.

5. If at least one of the coordinates y_1 and y_2 is zero, then $P(y_1, y_2) = 0$. This is a direct consequence of the previous observation.

Given all these considerations, the main theorem of this section is a simple consequence of Lemma 30.

Theorem 31 *In the setting described in Section 3.1, there is a number $\delta' > 0$, a neighborhood of the origin U', and a C^2-diffeomorphism $f : U' \to (-\delta', \delta')$ with inverse $g : (-\delta', \delta')^2 \to U'$ such that the following property holds.*

If $X_\epsilon(0) \in U$, then the stochastic process $Y_\epsilon = (Y_{\epsilon,1}, Y_{\epsilon,2})$ given by

$$Y_\epsilon(t) = f(X_\epsilon(t \wedge \tau_\epsilon^U))$$

satisfies the following system of SDEs up to τ_ϵ^U :

$$dY_{\epsilon,1} = (\lambda_+ Y_{\epsilon,1} + H_1(Y_\epsilon, \epsilon))\, dt + \epsilon \tilde{\sigma}_1(Y_\epsilon) dW \tag{2.16}$$
$$dY_{\epsilon,2} = (-\lambda_- Y_{\epsilon,2} + H_2(Y_\epsilon, \epsilon))\, dt + \epsilon \tilde{\sigma}_2(Y_\epsilon) dW, \tag{2.17}$$

where $\tilde{\sigma}_i : (-\delta', \delta')^2 \to \mathbb{R}$ are C^1 functions for $i = 1, 2$. The functions H_i are given by $H_i = \hat{H}_i + \epsilon^2 \Psi_i$, where $\Psi_i : (-\delta', \delta')^2 \to \mathbb{R}^2$ are continuous bounded functions, and $\hat{H}_i : (-\delta', \delta')^2 \times [0, \infty)$ are polynomials, so that for some constant $K_1 > 0$ and for any $y \in (-\delta', \delta')^2$,

$$|\hat{H}_1(y)| \le K_1 |y_1|^{\alpha_1^+} |y_2|^{\alpha_2^+},$$
$$|\hat{H}_2(y)| \le K_1 |y_1|^{\alpha_1^-} |y_2|^{\alpha_2^-}.$$

Here, the integer numbers $\alpha_i^\pm$, $i = 1, 2$, are such that (α_1^+, α_2^+) is of the form (2.14) *for some choice of $r = r_1 \ge 3$, and and (α_1^-, α_2^-) is of the form* (2.15) *for some choice $r = r_2 \ge 3$. In particular,*

$$|H_1(y, \epsilon)| \le K_1 y_1^2 |y_2| + K_2 \epsilon^2,$$
$$|H_2(y, \epsilon)| \le K_1 |y_1| y_2^2 + K_2 \epsilon^2,$$

for some constants $K_1 > 0$ and $K_2 > 0$.

2.4 Proof of Theorem 25

In this section we derive Theorem 25 from several auxiliary statements. Their proofs are postponed to later sections.

Theorem 31 allows to work with process $Y_\epsilon = f(X_\epsilon)$ instead of X_ϵ while Y_ϵ stays in $(-\delta', \delta')^2$

If we take $\delta \in (0, \delta')$, then for the initial conditions considered in Theorem 25 and given in (2.4),

$$\mathbf{P}\{X_\epsilon(0) \in U'\} \to 1, \quad \epsilon \to 0,$$

i.e.,

$$\mathbf{P}\{Y_\epsilon(0) \in (-\delta', \delta')^2\} \to 1, \quad \epsilon \to 0.$$

Moreover, denoting $f(x)$ by $y = (0, y_2)$ we can write

$$Y_\epsilon(0) = y + \epsilon^\alpha \chi_\epsilon = (\epsilon^\alpha \chi_{\epsilon,1},\ y_2 + \epsilon^\alpha \chi_{\epsilon,2}), \quad \epsilon > 0,$$

where $\chi_\epsilon = (\chi_{\epsilon,1}, \chi_{\epsilon,2})$ is a random vector convergent in distribution to $\chi_0 = (\chi_{0,1}, \chi_{0,2}) = \nabla f(x)\xi_0$. Due to the hypothesis in Theorem 25, we notice that the distribution of $\chi_{0,1}$ has no atom at 0.

Let us take any $p \in (0,1)$ such that

$$1 - \frac{\lambda_+}{\lambda_-} < p < \frac{\lambda_-}{\lambda_+ + \lambda_-}, \tag{2.18}$$

and define the following stopping time:

$$\hat{\tau}_\epsilon = \inf\{t : |Y_{\epsilon,1}(t)| = \epsilon^{\alpha p}\}.$$

Up to time $\hat{\tau}_\epsilon$, the process X_ϵ mostly evolves along the stable manifold $\mathcal{W}^s$. After $\hat{\tau}_\epsilon$, it evolves mostly along the unstable manifold $\mathcal{W}^u$. Process Y_ϵ evolves accordingly, along the images of $\mathcal{W}^s$ and $\mathcal{W}^u$ coinciding with the coordinate axes.

Let us introduce random variables $\eta_\epsilon^\pm$ via

$$\begin{aligned} \eta_\epsilon^+ &= \epsilon^{-\alpha} e^{-\lambda_+ \hat{\tau}_\epsilon} Y_{\epsilon,1}(\hat{\tau}_\epsilon), \\ \eta_\epsilon^- &= \epsilon^{-\alpha(1-p)\lambda_-/\lambda_+} Y_{\epsilon,2}(\hat{\tau}_\epsilon). \end{aligned}$$

Also we define the distribution of random vector (η_0^+, η_0^-) via

$$\begin{aligned} \eta_0^+ &= \chi_{0,1} + \mathbf{1}_{\{\alpha=1\}} N^+, \\ \eta_0^- &= |\eta_0^+|^{\lambda_-/\lambda_+} y_2, \end{aligned} \tag{2.19}$$

where

$$N^+ = \int_0^\infty e^{-\lambda_- s} \tilde{\sigma}_1(0, e^{-\lambda_- s} y_2) dW \tag{2.20}$$

is independent of $\chi_{0,1}$.

Lemma 32 *If the first inequality in* (2.18) *holds, then*

$$\mathbf{P}\{Y_{\epsilon,1}(\hat{\tau}_\epsilon) = \epsilon^{\alpha p} \operatorname{sgn} \eta_\epsilon^+\} \to 1, \quad \epsilon \to 0. \tag{2.21}$$

and

$$\left(\eta_\epsilon^+, \eta_\epsilon^-, \hat{\tau}_\epsilon + \frac{\alpha}{\lambda_+}(1-p)\log\epsilon\right) \xrightarrow{Law} \left(\eta_0^+, \eta_0^-, -\frac{1}{\lambda_+}\log|\eta_0^+|\right), \quad \epsilon \to 0. \tag{2.22}$$

We prove this lemma in Section 2.5. Along with the strong Markov property, it allows to reduce the study of the evolution of Y_ϵ after $\hat{\tau}_\epsilon$ to studying the solution of system (2.16)–(2.17) with initial condition

$$Y_\epsilon(0) = (\epsilon^{\alpha p}\operatorname{sgn}\eta_\epsilon^+, \epsilon^{\alpha(1-p)\lambda_-/\lambda_+}\eta_\epsilon^-), \tag{2.23}$$

where

$$(\eta_\epsilon^+, \eta_\epsilon^-) \xrightarrow{Law} (\eta_0^+, \eta_0^-), \quad \epsilon \to 0. \tag{2.24}$$

We denote

$$\tau_\epsilon = \tau_\epsilon(\delta) = \inf\{t \geq 0 : |Y_{\epsilon,1}(t)| = \delta\}. \tag{2.25}$$

Our next goal is to describe the behavior of $Y(\tau_\epsilon)$. To that end, we introduce a random variable θ via

$$\theta \overset{Law}{=} \begin{cases} N, & \alpha\lambda_- > \lambda_+, \\ \left(\frac{|\eta_0^+|}{\delta}\right)^{\lambda_-/\lambda_+} y_2 + N, & \alpha\lambda_- = \lambda_+, \\ \left(\frac{|\eta_0^+|}{\delta}\right)^{\lambda_-/\lambda_+} y_2, & \alpha\lambda_- < \lambda_+. \end{cases} \tag{2.26}$$

where the distribution of N conditioned on η_0^+, on $\{\operatorname{sgn}\eta_0^+ = \pm 1\}$ is centered Gaussian with variance

$$\sigma_\pm = \int_{-\infty}^0 e^{2\lambda_- s}\left|\tilde{\sigma}_2(\pm\delta e^{\lambda_+ s}, 0)\right|^2 ds.$$

Let us also recall that β is defined in (2.3).

Lemma 33 *Consider the solution to system* (2.16)–(2.17) *equipped with initial conditions* (2.23) *satisfying* (2.24). *If the second inequality in* (2.18) *holds, then*

$$\mathbf{P}\{|Y_{\epsilon,1}(\tau_\epsilon)| = \delta\} \to 1, \quad \epsilon \to 0, \tag{2.27}$$

$$\tau_\epsilon + \frac{\alpha p}{\lambda_+}\log\epsilon \xrightarrow{\mathbf{P}} \frac{1}{\lambda_+}\log\delta, \tag{2.28}$$

$$\epsilon^{-\beta} Y_{\epsilon,2}(\tau_\epsilon) \xrightarrow{Law} \theta. \tag{2.29}$$

Moreover, if $\beta < 1$, then the convergence in probability also holds.

A proof of this lemma is given in Section 2.6.

Now Theorem 25 follows from Lemmas 32 and 33. In fact, the strong Markov property and (2.21) imply

$$\mathbf{P}\{\tau_\epsilon^U = \hat{\tau}_\epsilon + \tau_\epsilon(\delta)\} \to 1, \quad \epsilon \to 0,$$

so that the asymptotics for τ_ϵ^U is defined by that of $\hat{\tau}_\epsilon$ and $\tau_\epsilon(\delta)$. It is also clear that one can set $\psi'_\epsilon = \eta_\epsilon^+$, and $\phi'_\epsilon = \nabla g(\mathrm{sgn}(\eta_\epsilon^+)\delta e_1)Y_\epsilon(\tau_\epsilon)$, so that the limiting distribution of Θ'_ϵ is given by

$$\left(\eta_0^+,\ \nabla g(\mathrm{sgn}(\eta_0^+)\delta e_1)(\theta e_2),\ \frac{1}{\lambda_+}\log\frac{\delta}{|\eta_0^+|}\right), \tag{2.30}$$

where random variables η_0^+ and θ are defined in (2.19) and (2.26)

2.5 Proof of Lemma 32

In this section we shall prove Lemma 32 using several auxiliary lemmas. We start with some terminology.

Definition 34 Given a family $(\xi_\epsilon)_{\epsilon>0}$ of random variables or random vectors and a function $h : (0,\infty) \to (0,\infty)$ we say that $\xi_\epsilon = O_{\mathbf{P}}(h(\epsilon))$ if for some $\epsilon_0 > 0$ distributions of $(\xi_\epsilon/h(\epsilon))_{0<\epsilon<\epsilon_0}$, form a tight family, i.e., for any $\delta > 0$ there is a constant $K_\delta > 0$ such that

$$\mathbf{P}\left\{|\xi_\epsilon| > K_\delta h(\epsilon)\right\} < \delta, \quad 0 < \epsilon < \epsilon_0.$$

Definition 35 A family of random variables or random vectors $(\xi_\epsilon)_{\epsilon>0}$ is called slowly growing as $\epsilon \to 0$ (or just slowly growing) if $\xi_\epsilon = O_{\mathbf{P}}(\epsilon^{-r})$ for all $r > 0$.

Our first lemma estimates the martingale component of the solution of SDEs (2.16) and (2.17). Let us define

$$\begin{aligned}
S_\epsilon^+(T) &= \sup_{t\le T}\left|\int_0^t e^{-\lambda_+ s}\tilde{\sigma}_1(Y_\epsilon(s))dW(s)\right|, \quad T > 0,\\
S_\epsilon^-(T) &= \sup_{t\le T}\left|\int_0^t e^{-\lambda_-(t-s)}\tilde{\sigma}_2(Y_\epsilon(s))dW(s)\right|, \quad T > 0.
\end{aligned}$$

Lemma 36 *Suppose $(\tau_\epsilon)_{\epsilon>0}$ is a family of stopping times (w.r.t. the natural filtration of W). Then*

$$S_\epsilon^+(\tau_\epsilon) = O_{\mathbf{P}}(1).$$

If additionally $(\tau_\epsilon)_{\epsilon>0}$ is slowly growing, then $S_\epsilon^-(\tau_\epsilon)$ is also slowly growing.

Proof. Let us start with the proof for S_ϵ^+. Use BDG inequality (see [41, Theorem 3.3.28]) and Itô's isometry to see that for every constant $K > 0$,

$$\begin{aligned}\mathbf{P}\left\{S_\epsilon^+(\tau_\epsilon) > K\right\} &\le \frac{1}{K^2}\mathbf{E}S_\epsilon^+(\tau_\epsilon)\\ &\le \frac{C_1}{K^2}\mathbf{E}\int_0^{\tau_\epsilon} e^{-2\lambda_+ s}\tilde{\sigma}_1(Y_\epsilon(s))ds.\end{aligned}$$

Since $Y_\epsilon(t) = f(X_\epsilon(t \wedge \tau_\epsilon^U))$, the process $t \mapsto \tilde{\sigma}_1(Y_\epsilon(t))$ is almost surely bounded. Hence, integrability of the exponential $t \mapsto e^{-2\lambda_+ t}$ implies that for any $\delta > 0$, there is a $K_\delta > 0$ such that

$$\sup_{\epsilon>0}\mathbf{P}\left\{S_\epsilon^+(\tau_\epsilon) > K_\delta\right\} \le \delta,$$

proving the first part of the lemma.

For the second part, fix $\delta > 0$ and $r > 0$. For every $0 < \rho < 2r$, there is $K_\rho > 0$ and $\epsilon_0 > 0$ such that

$$\sup_{0<\epsilon<\epsilon_0}\mathbf{P}\left\{\epsilon^\rho\tau_\epsilon > K_\rho\right\} < \delta/2.$$

Then, for an arbitrary $K > 0$, $0 < \epsilon < \epsilon_0$ and $0 < \rho < 2r$, it holds that

$$\begin{aligned}\mathbf{P}\left\{\epsilon^r S_\epsilon^-(\tau_\epsilon) > K\right\} &\le \mathbf{P}\left\{\tau_\epsilon > \epsilon^{-\rho}K_\rho\right\} + \mathbf{P}\left\{\epsilon^r S_\epsilon^-(\tau_\epsilon) > K, \tau_\epsilon \le \epsilon^{-\rho}K_\rho\right\}\\ &\le \delta/2 + \sum_{k=1}^{\lceil K_\rho\epsilon^{-\rho}\rceil}\mathbf{P}\left\{\epsilon^r\sup_{(k-1)\le t<k}\left|\int_0^t e^{-\lambda_-(t-s)}\tilde{\sigma}_2(Y_\epsilon(s))dW(s)\right| > K\right\}.\end{aligned}$$

In order to bound each probability in the last sum, proceed as for the other case:

$$\begin{aligned}&\mathbf{P}\left\{\epsilon^r\sup_{(k-1)\le t<k}\left|\int_0^t e^{-\lambda_-(t-s)}\tilde{\sigma}_2(Y_\epsilon(s))dW(s)\right| > K\right\}\\ &\le \mathbf{P}\left\{\epsilon^r e^{-(k-1)\lambda_-}\sup_{0\le t<k}\left|\int_0^t e^{\lambda_- s}\tilde{\sigma}_2(Y_\epsilon(s))dW(s)\right| > K\right\}\\ &\le \frac{\epsilon^{2r}e^{-2(k-1)\lambda_-}}{K^2}\mathbf{E}\int_0^k e^{2\lambda_- s}|\tilde{\sigma}_2(Y_\epsilon(s))|^2 ds\\ &\le \frac{\epsilon^{2r}C_2}{K^2},\end{aligned}$$

for some constant $C_2 > 0$. Hence, there is a constant $C_3 > 0$ such that

$$\mathbf{P}\left\{\epsilon^r S_\epsilon^-(\tau_\epsilon) > K\right\} \le \delta/2 + \frac{C_3}{K^2}\epsilon^{2r-\rho},$$

which implies the result and finishes the proof. ■

Lemma 37 *Suppose Y_ϵ is the solution of equations (2.16)–(2.17) with initial conditions given by*

$$Y_{\epsilon,1}(0) = \epsilon^\alpha \chi_{\epsilon,1} \quad \text{and} \quad Y_{\epsilon,2}(0) = y_2 + \epsilon^\alpha \chi_{\epsilon,2}, \tag{2.31}$$

where distributions of random variables $(\chi_{\epsilon,1})_{\epsilon>0}$ and $(\chi_{\epsilon,2})_{\epsilon>0}$ form tight families. Let us fix any $R > 0$ and denote $l_\epsilon = \tau_\epsilon^U \wedge (-\frac{\alpha}{\lambda_+}\log\epsilon + R)$ for $\epsilon > 0$. Then

$$\sup_{t\le l_\epsilon} e^{-\lambda t}|Y_{\epsilon,1}(t)| = O_{\mathbf{P}}(\epsilon^\alpha),$$

and the family

$$\left(\epsilon^{-\alpha}\sup_{t\le l_\epsilon}|Y_{\epsilon,2}(t) - e^{-\lambda t}(y_2 + \epsilon^\alpha\chi_{\epsilon,2})|\right)_{\epsilon>0}$$

is slowly growing.

Proof. The tightness property implies that without loss of generality we can assume that $|\chi_{\epsilon,1}|, |\chi_{\epsilon,2}| < C$ for some constant $C > 0$ and every $\epsilon > 0$.

Let us fix $\gamma > 0$. We can use Lemma 36 to take $c = c(\gamma/3) > 0$ such that

$$\mathbf{P}\{S_\epsilon^+(l_\epsilon) > c\} < \gamma/2,$$

and

$$\mathbf{P}\{S_\epsilon^-(l_\epsilon) > c\epsilon^{-q}\} < \gamma/2,$$

where q is an arbitrary number satisfying $0 < q < \alpha$. Let us introduce a constant $K = (3c) \vee C$ and stopping times

$$\begin{aligned}
\beta_+ &= \inf\left\{t \ge 0 : e^{-\lambda_+ t}|Y_{\epsilon,1}(t)| \ge 2K\epsilon^\alpha\right\},\\
\beta_- &= \inf\left\{t \ge 0 : |Y_{\epsilon,2}(t) - e^{-\lambda_- t}(y_2 + \epsilon^\alpha\chi_{\epsilon,2})| \ge 2K\epsilon^{\alpha-q}\right\},\\
\beta &= \beta_+ \wedge \beta_- \wedge l_\epsilon.
\end{aligned}$$

We start with an estimate for $Y_{\epsilon,1}$. Duhamel's principle for (2.16), Theorem 31 and Lemma 36 imply that the estimate

$$\begin{aligned}
\sup_{t\le\beta} e^{-\lambda_+ t}|Y_{\epsilon,1}(t)| &\le \epsilon^\alpha K + K_1\int_0^\beta e^{-\lambda_+ s}Y_{\epsilon,1}(s)^2|Y_{\epsilon,2}(s)|ds + K_2\frac{\epsilon^2}{\lambda_+} + \epsilon S_\epsilon^+(\beta)\\
&\le \epsilon^\alpha K + K_1\int_0^\beta e^{-\lambda_+ s}Y_{\epsilon,1}(s)^2|Y_{\epsilon,2}(s)|ds + K_2\frac{\epsilon^2}{\lambda_+} + \epsilon\frac{K}{3}
\end{aligned} \tag{2.32}$$

holds with probability at least $1-\gamma/2$. We analyze each term in the RHS of equation (2.32).

Let us start with the integral in (2.32). For $s \le \beta$, we see that

$$\begin{aligned} Y_{\epsilon,1}(s)^2|Y_{\epsilon,2}(s)| &\le 4K^2\epsilon^{2\alpha}e^{2\lambda_+ s}\Big(|Y_{\epsilon,2}(s)-e^{-\lambda_- s}(y_2+\epsilon^\alpha\chi_{\epsilon,2})|+e^{-\lambda_- s}|y_2+\epsilon^\alpha\chi_{\epsilon,2}|\Big) \\ &\le 8K^3\epsilon^{3\alpha-q}e^{2\lambda_+ s}+4K^2\epsilon^{2\alpha}e^{(2\lambda_+-\lambda_-)s}(|y_2|+\epsilon^\alpha C). \end{aligned}$$

Therefore,

$$\begin{aligned} K_1\int_0^\beta e^{-\lambda_+ s}Y_{\epsilon,1}(s)^2|Y_{\epsilon,1}(s)|ds &\le \frac{8K^3K_1e^{\lambda_+R}}{\lambda_+}\epsilon^{2\alpha-q} \\ &\quad + 4K_1K^2\epsilon^{2\alpha}(|y_2|+\epsilon^\alpha C)\int_0^\beta e^{(\lambda_+-\lambda_-)s}ds \\ &\le K\epsilon^\alpha/12+5K_1K^2\epsilon^{2\alpha}|y_2|\int_0^\beta e^{(\lambda_+-\lambda_-)s}ds \end{aligned} \tag{2.33}$$

for all $\epsilon>0$ small enough. Notice that this is a rough estimate, the constants on the r.h.s. are not optimal but sufficient for our purposes. This also applies to some other estimates in this proof.

Let us estimate the integral on the r.h.s. of (2.33). When $\lambda_+>\lambda_-$, the integral is bounded by

$$\frac{1}{\lambda_+-\lambda_-}e^{(\lambda_+-\lambda_-)\beta} \le \frac{e^{(\lambda_+-\lambda_-)R}}{\lambda_+-\lambda_-}\epsilon^{-\alpha+\alpha\lambda_-/\lambda_+};$$

if $\lambda_+<\lambda_-$, then the integral on the r.h.s of (2.33) is bounded by $(\lambda_--\lambda_+)^{-1}$; if $\lambda_+=\lambda_-$, then the integral is bounded by $2\alpha\lambda_+^{-1}|\log\epsilon|$. Hence, for some constant $K_{\lambda_+,\lambda_-}>0$ and $\epsilon>0$ small enough,

$$\begin{aligned} K_1\int_0^\beta e^{-\lambda_+ s}Y_{\epsilon,1}(s)^2|Y_{\epsilon,2}(s)|ds &\le K\epsilon^\alpha/12+K_{\lambda_+,\lambda_-}\epsilon^{2\alpha-\alpha(1-\lambda_-/\lambda_+)^+}|\log\epsilon| \\ &\le K\epsilon^\alpha/6. \end{aligned} \tag{2.34}$$

Also, for $\epsilon>0$ small enough,

$$K_2\epsilon^2/\lambda_+ + \epsilon K/3 < K\epsilon^\alpha/2. \tag{2.35}$$

From (2.32), (2.34) and (2.35) we get that for all $\epsilon>0$ small enough, the event

$$A=\left\{\sup_{t\le\beta}e^{-\lambda_+ t}|Y_{\epsilon,1}(t)|\le 5K\epsilon^\alpha/3\right\}$$

is such that $\mathbf{P}(A) > 1 - \gamma/2$.

Let us now consider $Y_{\epsilon,2}(t)$ and denote

$$Z_\epsilon(t) = Y_{\epsilon,2}(t) - e^{-\lambda_- t}(y_2 + \epsilon^\alpha \chi_{\epsilon,2}).$$

Duhamel's principle for $Y_{\epsilon,2}$, the definition of β, Theorem 31 and Lemma 36 imply that the inequalities

$$\begin{aligned}
\sup_{t\le\beta} |Z_\epsilon(t)| &\le K_1 \sup_{t\le\beta} \int_0^t e^{-\lambda_-(t-s)} |Y_{\epsilon,1}(s)|^{\alpha_1^-} |Y_{\epsilon,2}(s)|^{\alpha_2^-}\, ds + K_2\epsilon^2/\lambda_- + \epsilon S_\epsilon^-(\beta) \\
&\le K_1 \sup_{t\le\beta} \int_0^t e^{-\lambda_-(t-s)} |Y_{\epsilon,1}(s)|^{\alpha_1^-} |Y_{\epsilon,2}(s)|^{\alpha_2^-}\, ds \\
&\qquad\qquad + \epsilon^{\alpha-q}\left(K_2 \epsilon^{2-\alpha+q}/\lambda_- + \epsilon^{1-\alpha+q} S_\epsilon^-(\beta)\right) \\
&\le 2^{\alpha_1^-} \epsilon^{\alpha\alpha_1^-} K^{\alpha_1^-} K_1 \sup_{t\le\beta} e^{-\lambda_- t} \int_0^t e^{(\lambda_- + \alpha_1^-\lambda_+)s} |Y_{\epsilon,2}(s)|^{\alpha_2^-}\, ds + \epsilon^{\alpha-q} K/2
\end{aligned} \tag{2.36}$$

hold with probability at least $1-\gamma/2$ and for all $\epsilon > 0$ small enough. We analyze the integral term in (2.36). Note that, from the definition of β, and the inequality $(a+b)^r \le 2^{r-1}(a^r + b^r)$ we have that for any $t \le \beta$ and any $\epsilon > 0$ small enough,

$$\begin{aligned}
|Y_{\epsilon,2}(t)|^{\alpha_2^-} &\le 2^{\alpha_2^- - 1} Z_\epsilon(t)^{\alpha_2^-} + 2^{\alpha_2^- - 1} e^{-\alpha_2^-\lambda_- t} |y_2 + \epsilon^\alpha \chi_{\epsilon,2}|^{\alpha_2^-} \\
&\le 2^{2\alpha_2^- - 1} K^{\alpha_2^-} \epsilon^{(\alpha-q)\alpha_2^-} + 2^{2(\alpha_2^- - 1)} e^{-\alpha_2^-\lambda_- t} |y_2|^{\alpha_2^-} \\
&\qquad\qquad + 2^{2(\alpha_2^- - 1)} \epsilon^{\alpha\alpha_2^-} e^{-\alpha_2^-\lambda_- t} |\chi_{\epsilon,2}|^{\alpha_2^-} \\
&\le \epsilon^{\alpha_2^-(\alpha-q)} 2^{2(\alpha_2^- - 1)} \left(2K^{\alpha_2^-} + \epsilon^{q\alpha_2^-} |\chi_{\epsilon,2}|^{\alpha_2^-}\right) + 2^{2(\alpha_2^- - 1)} e^{-\alpha_2^-\lambda_- t} |y_2|^{\alpha_2^-}.
\end{aligned}$$

Hence there is a constant $K_\alpha > 0$ such that

$$|Y_{\epsilon,2}(t)|^{\alpha_2^-} \le \epsilon^{\alpha_2^-(\alpha-q)} K_\alpha + K_\alpha e^{-\alpha_2^-\lambda_- t}, \quad t \le \beta.$$

Using the last inequality, the definition of β, and the fact $\alpha_1^-\lambda_+ - (\alpha_2^- - 1)\lambda_- = 0$ from Theorem 31, we get

$$\begin{aligned}
&\epsilon^{\alpha\alpha_1^-} e^{-\lambda_- t} \int_0^t e^{(\lambda_- + \alpha_1^-\lambda_+)s} |Y_{\epsilon,2}(s)|^{\alpha_1^-}\, ds \\
&\qquad \le \epsilon^{\alpha(\alpha_1^- + \alpha_2^-)} e^{\lambda_+\alpha_1^-\beta} \frac{K_\alpha \epsilon^{-q\alpha_2^-}}{\lambda_- + \alpha_1^-\lambda_+} + K_\alpha \epsilon^{\alpha\alpha_1^-} \int_0^t e^{(\alpha_1^-\lambda_+ - (\alpha_2^- - 1)\lambda_-)s} ds \\
&\qquad \le \epsilon^{(\alpha-q)\alpha_2^-} \frac{K_\alpha e^{\lambda_+\alpha_1^- R}}{\lambda_- + \alpha_1^-\lambda_+} + K_\alpha \epsilon^{\alpha\alpha_1^-} \beta.
\end{aligned} \tag{2.37}$$

Again, from Theorem 31 we know that $\alpha_1^- \geq 1$ and $\alpha_2^- \geq 2$ which together with (2.37) imply that for all $\epsilon > 0$ small enough

$$2^{\alpha_1^-}\epsilon^{\alpha\alpha_1^-}K^{\alpha_1^-}K_1 \sup_{t\leq\beta} e^{-\lambda_- t}\int_0^t e^{(\lambda_- + \alpha_1^-\lambda_+)s}|Y_{\epsilon,2}(s)|^{\alpha_2^-}ds \leq K\epsilon^{\alpha-q}/6. \quad (2.38)$$

Using (2.38) and (2.36) we conclude that the event

$$B = \left\{\sup_{t\leq\beta}|Y_{\epsilon,2}(t) - e^{-\lambda_- t}(y_2 + \epsilon^\alpha\chi_{\epsilon,2})| \leq 2K\epsilon^{\alpha-q}/3\right\}$$

is such that $\mathbf{P}(B) \geq 1 - \gamma/2$, for all $\epsilon > 0$ small enough.

The proof will be complete once we show that $\beta = l_\epsilon$ with probability at least $1-\gamma$. The latter is a consequence of the following chain of inequalities that hold for all $\epsilon > 0$ small enough:

$$\begin{aligned}\mathbf{P}\{\beta_+ \wedge \beta_- \leq l_\epsilon\} &\leq \mathbf{P}\left(\{\beta_+ \wedge \beta_- \leq l_\epsilon\} \cap A \cap B\right) + \mathbf{P}(A^c) + \mathbf{P}(B^c)\\ &\leq \mathbf{P}\left(\{\beta_+ \wedge \beta_- \leq l_\epsilon\} \cap A \cap B\right) + \gamma\\ &\leq \mathbf{P}\left(\{\beta_+ \leq \beta_- \wedge l_\epsilon\} \cap A\right) + \mathbf{P}\left(\{\beta_- \leq \beta_+ \wedge l_\epsilon\} \cap B\right) + \gamma\\ &= \mathbf{P}\{2 \leq 5/3\} + \mathbf{P}\{2 \leq 2/3\} + \gamma = \gamma.\end{aligned}$$

■

Let us now analyze the evolution of the process Y_ϵ up to time $\hat{\tau}_\epsilon \wedge \tau_\epsilon^U$. We start with an application of Duhamel's principle:

$$Y_{\epsilon,1}(t) = e^{\lambda_+ t}Y_{\epsilon,1}(0) + \int_0^t e^{\lambda_+(t-s)}H_1(Y_\epsilon(s),\epsilon)ds + \epsilon e^{\lambda_+ t}N_\epsilon^+(t), \quad (2.39)$$

$$Y_{\epsilon,2}(t) = e^{-\lambda_- t}Y_{\epsilon,2}(0) + \int_0^t e^{-\lambda_-(t-s)}H_2(Y_\epsilon(s),\epsilon)ds + \epsilon N_\epsilon^-(t), \quad (2.40)$$

where $N_\epsilon^\pm(t)$ are defined by

$$\begin{aligned}N_\epsilon^+(t) &= \int_0^t e^{-\lambda_+ s}\tilde{\sigma}_1(Y_\epsilon(s))dW(s),\\ N_\epsilon^-(t) &= \int_0^t e^{-\lambda_-(t-s)}\tilde{\sigma}_2(Y_\epsilon(s))dW(s). \quad (2.41)\end{aligned}$$

Lemma 38

$$\sup_{t\leq\hat{\tau}_\epsilon}|Y_{\epsilon,2}(t) - e^{-\lambda_- t}y_2| = O_{\mathbf{P}}(\epsilon^{\alpha p}).$$

Proof. Duhamel's principle, Theorem 31, and the definition of $\hat{\tau}_\epsilon$ imply that for some $K > 0$,

$$|Y_{\epsilon,2}(t) - e^{-\lambda_- t}y_2| \le \epsilon^\alpha |\chi_{\epsilon,2}| + \int_0^t e^{-\lambda_-(t-s)} \left(K_1|Y_{\epsilon,1}(s)|Y^2_{\epsilon,2}(s) + K_2\epsilon^2\right) ds + \epsilon S^-_\epsilon(t)$$
$$\le \epsilon^\alpha |\chi_{\epsilon,2}| + K\epsilon^{\alpha p} + \epsilon^{\alpha p}\left(\epsilon^{1-\alpha p} S^-_\epsilon(\hat{\tau}_\epsilon)\right)$$

for any $t \in (0, \hat{\tau}_\epsilon)$. The result follows since by Lemma 36 the r.h.s. is $O_{\mathbf{P}}(\epsilon^{\alpha p})$ ■

As a simple corollary of this lemma, the first statement in Theorem 32 follows:

Corollary 39 *As $\epsilon \to 0$,*

$$\mathbf{P}\{\tau^U_\epsilon < \hat{\tau}_\epsilon\} \to 0.$$

In particular, (2.21) *holds true.*

Lemma 40 *Let*

$$N^+_0(t) = \int_0^t e^{-\lambda_- s}\tilde{\sigma}_1(0, e^{-\lambda_- s}y_2)dW.$$

Then

$$\sup_{t\le\hat{\tau}_\epsilon} |N^+_\epsilon(t) - N^+_0(t)| \xrightarrow{L^2} 0, \quad \epsilon \to 0.$$

Proof. BDG inequality implies that for some constants $C_1, C_2 > 0$,

$$\mathbf{E}\sup_{t\le\hat{\tau}_\epsilon} |N^+_\epsilon(t) - N^+_0(t)|^2 \le C_1\mathbf{E}\int_0^{\hat{\tau}_\epsilon} e^{-2\lambda_+ s}|\tilde{\sigma}_1(Y_{\epsilon,1}(s), Y_{\epsilon,2}(s)) - (0, e^{-\lambda_- s}y_2)|^2 ds$$
$$\le C_2\mathbf{E}\sup_{t\le\hat{\tau}_\epsilon} |\tilde{\sigma}_1(Y_{\epsilon,1}(s), Y_{\epsilon,2}(s)) - \tilde{\sigma}_1(0, e^{-\lambda_- s}y_2)|^2. \tag{2.42}$$

From Lemma 38 and the definition of $\hat{\tau}_\epsilon$, it follows that

$$\sup_{t\le\hat{\tau}_\epsilon}\left|(Y_{\epsilon,1}(t), Y_{\epsilon,2}(t)) - (0, e^{-\lambda_- t}y_2)\right| = O_{\mathbf{P}}(\epsilon^{\alpha p}). \tag{2.43}$$

The desired convergence follows now from (2.42), (2.43), and the boundedness and Lipschitzness of $\tilde{\sigma}_1$. ■

We are now in position to give the first rough asymptotics for the time $\hat{\tau}_\epsilon$. From now on we restrict ourselves to the event $\{\tau^U_\epsilon > \hat{\tau}_\epsilon\}$ since due to Corollary 39 its probability is arbitrarily high.

Lemma 41 *As $\epsilon \to 0$,*

$$\mathbf{P}\left\{\hat{\tau}_\epsilon > -\frac{\alpha}{\lambda_+}\log\epsilon\right\} \to 0.$$

Proof. Let u_ϵ be the solution to the following SDE:

$$\begin{aligned} du_\epsilon(t) &= \lambda_+ u_\epsilon(t)dt + \epsilon\tilde{\sigma}_1(Y_\epsilon(t))dW(t), \\ u_\epsilon(0) &= \epsilon^\alpha \chi_{\epsilon,1}. \end{aligned}$$

Let us take $\delta_0 \in (0,1)$ to be specified later and consider the following stopping time

$$\widetilde{\tau}_\epsilon = \inf\left\{t : |u_\epsilon(t)| = \epsilon^{\alpha\delta_0}\right\}.$$

Duhamel's principle for u_ϵ writes as

$$\begin{aligned} u_\epsilon(t) &= \epsilon^\alpha e^{\lambda_+ t}\chi_{\epsilon,1} + \epsilon e^{\lambda_+ t} N_\epsilon^+(t) \\ &= \epsilon^\alpha e^{\lambda_+ t}\widetilde{\eta}_\epsilon(t), \end{aligned}$$

with

$$\widetilde{\eta}_\epsilon(t) = \chi_{\epsilon,1} + \epsilon^{1-\alpha}N_\epsilon^+(t). \tag{2.44}$$

Hence, the definition of $\widetilde{\tau}_\epsilon$ implies $\epsilon^{\alpha\delta_0} = \epsilon^\alpha e^{\lambda_+\widetilde{\tau}_\epsilon}|\widetilde{\eta}_\epsilon(\widetilde{\tau}_\epsilon)|$, so that

$$\widetilde{\tau}_\epsilon = -\frac{\alpha}{\lambda_+}(1-\delta_0)\log\epsilon - \frac{1}{\lambda_+}\log|\widetilde{\eta}_\epsilon(\widetilde{\tau}_\epsilon)|.$$

Due to (2.44) and Lemma 40, the distributions of $\frac{1}{\lambda_+}\log|\widetilde{\eta}_\epsilon(\widetilde{\tau}_\epsilon)|$ form a tight family. Therefore,

$$\lim_{\epsilon\to 0}\mathbf{P}\left\{\widetilde{\tau}_\epsilon > -(1-\delta_0^2)\frac{\alpha}{\lambda_+}\log\epsilon\right\} = 0. \tag{2.45}$$

This fact allows us to use Lemma 37 to estimate Y_ϵ up to $\hat{\tau}_\epsilon \wedge \widetilde{\tau}_\epsilon$. From (2.39), the difference $\Delta_\epsilon = Y_{\epsilon,1} - u_\epsilon$ is given by

$$\Delta_\epsilon(t) = e^{\lambda_+ t}\int_0^t e^{-\lambda_+ s}H_1(Y_\epsilon(s),\epsilon)ds.$$

We can use (2.45) to justify the application of Lemma 37 up to time $\hat{\tau}_\epsilon \wedge \widetilde{\tau}_\epsilon$. Then, we combine Theorem 31, Lemma 37, and the definition of $\hat{\tau}_\epsilon$ to see that

$$\begin{aligned} \sup_{t\le\hat{\tau}_\epsilon\wedge\widetilde{\tau}_\epsilon} e^{-\lambda_+ t}|H_1(Y_\epsilon(t),\epsilon)| &\le K_1 \sup_{t\le\hat{\tau}_\epsilon\wedge\widetilde{\tau}_\epsilon}\left(\left(e^{-\lambda_+ t}|Y_{\epsilon,1}(t)|\right)|Y_{\epsilon,1}(t)|\cdot|Y_{\epsilon,2}(t)|\right) + K_2\epsilon^2 \\ &= O_{\mathbf{P}}\left(\epsilon^{\alpha+\alpha p}\right) \end{aligned}$$

and

$$e^{\lambda_+ \hat{\tau}_\epsilon \wedge \widetilde{\tau}_\epsilon} = O_{\mathbf{P}}\left(\epsilon^{-\alpha(1-\delta_0^2)}\right).$$

These two estimates together with (2.45) imply

$$\sup_{t \le \hat{\tau}_\epsilon \wedge \widetilde{\tau}_\epsilon} |\Delta_\epsilon(t)| = O_{\mathbf{P}}\left(\epsilon^{\alpha(p+\delta_0^2)}|\log\epsilon|\right).$$

On one hand, (2.45) implies

$$\mathbf{P}\left(\left\{\hat{\tau}_\epsilon > -\frac{\alpha}{\lambda_+}\log\epsilon\right\} \cap \{\hat{\tau}_\epsilon \le \widetilde{\tau}_\epsilon\}\right) \to 0.$$

On the other hand, if $\hat{\tau}_\epsilon > \widetilde{\tau}_\epsilon$ then

$$|Y_{\epsilon,1}(\widetilde{\tau}_\epsilon)| = \left|\epsilon^{\alpha\delta_0} + O_{\mathbf{P}}(\epsilon^{\alpha(p+\delta_0^2)}|\log\epsilon|)\right|,$$

and

$$|Y_{\epsilon,1}(\widetilde{\tau}_\epsilon)| < \epsilon^{\alpha p}.$$

These relations contradict each other for sufficiently small ϵ if we choose $\delta_0 < p$. So, this choice of δ_0 guarantees that $\mathbf{P}\{\hat{\tau}_\epsilon > \widetilde{\tau}_\epsilon\} \to 0$ implying the result. ■

Proof of Lemma 32. Recall that we work on the high probability event $\{\hat{\tau}_\epsilon < \tau_\epsilon^U\}$. Hence, for each $\epsilon > 0$, we have the identity

$$\epsilon^{\alpha p} = \epsilon^\alpha e^{\lambda_+ \hat{\tau}_\epsilon}|\eta_\epsilon^+|.$$

Solving for $\hat{\tau}_\epsilon$ and then plugging it back into $Y_{\epsilon,1}$, we get

$$\begin{aligned} \hat{\tau}_\epsilon &= -\frac{\alpha}{\lambda_+}(1-p)\log\epsilon - \frac{1}{\lambda_+}\log|\eta_\epsilon^+|, \\ Y_{\epsilon,1}(\hat{\tau}_\epsilon) &= \epsilon^{\alpha p}\operatorname{sgn}(\eta_\epsilon^+). \end{aligned} \tag{2.46}$$

Using this information we are in position to get the asymptotic behavior of the random variables $\eta_\epsilon^\pm$. First, from relation (2.39) we get

$$\eta_\epsilon^+ = \chi_{\epsilon,1} + \epsilon^{-\alpha}\int_0^{\hat{\tau}_\epsilon} e^{-\lambda_+ s} H_1(Y_\epsilon(s),\epsilon)ds + \epsilon^{1-\alpha}N_\epsilon^+(\hat{\tau}_\epsilon). \tag{2.47}$$

Using (2.46) in (2.40) we get

$$\begin{aligned} \eta_\epsilon^- &= |\eta_\epsilon^+|^{\lambda_-/\lambda_+}(y_2 + \epsilon^\alpha\chi_{\epsilon,2}) + |\eta_\epsilon^+|^{\lambda_-/\lambda_+}\int_0^{\hat{\tau}_\epsilon} e^{\lambda_- s}H_2(Y_\epsilon(s),\epsilon)ds \\ &\quad + \epsilon^{1-\alpha(1-p)\lambda_-/\lambda_+}N_\epsilon^-(\hat{\tau}_\epsilon). \end{aligned} \tag{2.48}$$

The main part of the proof is based on representations (2.46)–(2.48).

Lemma 41 allows us to use the estimates established in Lemma 37 up to time $\hat{\tau}_\epsilon$. In particular, now we can conclude that the family

$$\left(\epsilon^{-\alpha} \sup_{t \le \hat{\tau}_\epsilon} |Y_{\epsilon,2}(t) - e^{-\lambda_- t} y_2|\right)_{\epsilon>0} \tag{2.49}$$

is slowly growing thus improving Lemma 38.

To obtain the desired convergence for η_ϵ^+, we analyze the r.h.s. of (2.47) term by term. The convergence of the first term was one of our assumptions. For the second one, we need to estimate $H_1(Y_\epsilon, \epsilon)$. Using Lemma 37, the boundness of $Y_{\epsilon,2}$ and the definition of $\hat{\tau}_\epsilon$, we see that

$$\sup_{t \le \hat{\tau}_\epsilon} e^{-\lambda_+ t} Y_{\epsilon,1}^2(t) |Y_{\epsilon,2}(t)| = O_{\mathbf{P}}(\epsilon^{\alpha+\alpha p}). \tag{2.50}$$

This estimate and Theorem 31 imply that

$$\begin{aligned} \epsilon^{-\alpha} \int_0^{\hat{\tau}_\epsilon} e^{-\lambda_+ s} H_1(Y_\epsilon(s), \epsilon) ds &\le K_1 \epsilon^{-\alpha} \int_0^{\hat{\tau}_\epsilon} e^{-\lambda_+ s} Y_{\epsilon,1}^2(s) |Y_{\epsilon,2}(s)| ds + \frac{K_2}{\lambda_+} \epsilon^{2-\alpha} \\ &= O_{\mathbf{P}}(\epsilon^{\alpha p} |\log \epsilon|). \end{aligned}$$

Let us estimate the third term in (2.47). We can use the last estimate along with (2.47) and Lemma 40 to conclude that the distributions of positive part of $\lambda_+^{-1} \log |\eta_\epsilon^+|$ form a tight family. Therefore, (2.46) implies that

$$\hat{\tau}_\epsilon \xrightarrow{\mathbf{P}} \infty, \quad \epsilon \to 0.$$

Combined with Itô isometry and Lemma 40, this implies

$$N_\epsilon^+(\hat{\tau}_\epsilon) \xrightarrow{L^2} N^+, \quad \epsilon \to 0,$$

which completes the analysis of η_ϵ^+ and, due to (2.46), of $\hat{\tau}_\epsilon$.

To obtain the convergence of η_ϵ^-, we study (2.48). Combining (2.49), the inequality

$$|Y_{\epsilon,1}(t)| Y_{\epsilon,2}^2(t) \le 2|Y_{\epsilon,1}(t)| \left(|Y_{\epsilon,2}(t) - e^{-\lambda_- t} y_2|^2 + e^{-2\lambda_- t} y_2^2\right),$$

and the definition of $\hat{\tau}_\epsilon$ we see that for any $q \in (0, \alpha p)$,

$$\sup_{t \le \hat{\tau}_\epsilon} e^{\lambda_- t} |Y_{\epsilon,1}(t)| Y_{\epsilon,2}^2(t) = O_{\mathbf{P}}\left(\epsilon^{\alpha p + \alpha - q} e^{\lambda_- \hat{\tau}_\epsilon} + \epsilon^{\alpha p}\right).$$

Hence, as a consequence of Theorem 31 and (2.46) we have

$$\int_0^{\hat{\tau}_\epsilon} e^{\lambda_- s} H_2(Y_\epsilon(s),\epsilon)ds = O_{\mathbf{P}}\left(\left(\epsilon^{\alpha p-q+\alpha}e^{\lambda_-\hat{\tau}_\epsilon}+\epsilon^{\alpha p}\right)|\log\epsilon|\right)$$
$$= O_{\mathbf{P}}\left(\left(\epsilon^{\alpha(1-(1-p)\lambda_-/\lambda_+)+(\alpha p-q)}+\epsilon^{\alpha p}\right)|\log\epsilon|\right).$$

Combining this and Lemma 36 in (2.48) we obtain

$$\eta_\epsilon^- = |\eta_\epsilon^+|^{\lambda_-/\lambda_+}y_2 + O_{\mathbf{P}}(\epsilon^\alpha) + O_{\mathbf{P}}\left(\left(\epsilon^{\alpha(1-(1-p)\lambda_-/\lambda_+)+(\alpha p-q)}+\epsilon^{\alpha p}\right)|\log\epsilon|\right)$$
$$+ O_{\mathbf{P}}\left(\epsilon^{1-\alpha(1-p)\lambda_-/\lambda_+-q}\right)$$

which finishes the proof of Lemma 32 by choosing q small enough. ∎

2.6 Proof of Lemma 33

Consider the solution to system (2.16)–(2.17) equipped with initial conditions (2.23) satisfying (2.24). Let us restrict the analysis to the arbitrary high probability event

$$\{|\eta_\epsilon^\pm| \le K_\pm\},$$

for some constants $K_\pm > 0$.

Lemma 42 *Let $p\in(0,1)$ satisfy (2.18), and let $(t_\epsilon)_{\epsilon>0}$ be a slowly growing family of stopping times. Consider $t'_\epsilon = t_\epsilon \wedge \tau_\epsilon^U$, then for any $\gamma>0$,*

$$\lim_{\epsilon\to 0}\mathbf{P}\left\{\sup_{t\le t'_\epsilon}|Y_{\epsilon,2}(t)| \le (K_-+\gamma)\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}\right\} = 1.$$

Proof. Let $\gamma>0$. We recall that N_ϵ^- is defined in (2.41) and introduce the process

$$M_\epsilon(t) = N_\epsilon^-(t) + \epsilon\int_0^t e^{-\lambda_-(t-s)}\Psi_2(Y_\epsilon(s))ds, \tag{2.51}$$

where Ψ_2 was introduced in Theorem 31, and the stopping time

$$\beta_\epsilon = \inf\left\{t : |Y_{\epsilon,2}(t)| > (K_-+\gamma)\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}\right\}.$$

Using the fact that $Y_{\epsilon,1}$ is bounded, it is easy to see that there is a constant K_{λ_-} independent of t, so that for any $t\le\beta_\epsilon\wedge t'_\epsilon$, we have

$$\int_0^t e^{-\lambda_-(t-s)}|Y_{\epsilon,1}(s)|Y_{\epsilon,2}^2(s)ds \le K_{\lambda_-}\epsilon^{2\alpha(1-p)\lambda_-/\lambda_+}.$$

This estimate, along with Duhamel's principle and Theorem 31 implies that for some constant $C > 0$ and any $t \leq \beta_\epsilon \wedge t'_\epsilon$,

$$\begin{aligned}|Y_{\epsilon,2}(t)| &\leq \epsilon^{\alpha(1-p)\lambda_-/\lambda_+}|\eta^-_\epsilon| + K_1 \int_0^t e^{-\lambda_-(t-s)}|Y_{\epsilon,1}(s)|Y^2_{\epsilon,2}(s)ds + \epsilon \sup_{t\leq\beta_\epsilon}|M_\epsilon(t)| \\ &\leq \epsilon^{\alpha(1-p)\lambda_-/\lambda_+}K_- + C\epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} + \epsilon \sup_{t\leq\beta_\epsilon}|M_\epsilon(t)|.\end{aligned}$$

Hence, using Lemma 36 to estimate M_ϵ, we obtain that

$$\begin{aligned}\mathbf{P}\{\beta_\epsilon < t'_\epsilon\} &= \mathbf{P}\left\{\sup_{t\leq\beta_\epsilon\wedge t'_\epsilon}|Y_{\epsilon,2}(t)| \geq (K_- + \gamma)\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}\right\} \\ &\leq \mathbf{P}\left\{C\epsilon^{\alpha(1-p)\lambda_-/\lambda_+} + \epsilon^{1-\alpha(1-p)\lambda_-/\lambda_+}\sup_{t\leq\beta_\epsilon}|M_\epsilon(t)| \geq \gamma\right\}\end{aligned}$$

converges to 0 as $\epsilon \to 0$ proving the lemma. ∎

Lemma 43 *Under the assumptions of lemma 42, for any $\rho \in (0, \frac{\alpha p}{\lambda_+}]$, $\gamma > 0$, and $C > 0$, define $\rho_\epsilon = (-\rho\log\epsilon + C) \wedge \tau^U_\epsilon$. Then, we have*

$$\lim_{\epsilon\to 0}\mathbf{P}\left\{\sup_{t\leq\rho_\epsilon}|Y_{\epsilon,1}(t)|e^{-\lambda_+ t} \leq (1+\gamma)\epsilon^{\alpha p}\right\} = 1.$$

Proof. Define the stopping time

$$\beta_\epsilon = \inf\left\{t : |Y_{\epsilon,1}(t)|e^{-\lambda_+ t} \geq (1+\gamma)\epsilon^{\alpha p}\right\}.$$

As a consequence of Duhamel's principle and Theorem 31 we get the bound

$$\begin{aligned}\sup_{t\leq\beta_\epsilon\wedge\rho_\epsilon}|Y_{\epsilon,1}(t)|e^{-\lambda_+ t} \leq& \epsilon^{\alpha p} + K_1\int_0^{\beta_\epsilon\wedge\rho_\epsilon} e^{-\lambda_+ s}Y^2_{\epsilon,1}(s)|Y_{\epsilon,2}(s)|ds \\ &+ \epsilon^2 K_2\lambda_+^{-1} + \epsilon S^+_\epsilon(\beta_\epsilon).\end{aligned}$$

This estimate together with Lemma 42, Lemma 36 and the definition of ρ_ϵ implies that for any small $\delta > 0$ we can find a constant $K > 0$, so that with probability bigger than $1-\delta$, the inequalities

$$\begin{aligned}\sup_{t\leq\beta_\epsilon\wedge\rho_\epsilon}|Y_{\epsilon,1}(t)|e^{-\lambda_+ t} &\leq \epsilon^{\alpha p} + K\epsilon^{\alpha p+\alpha(1-p)\lambda_-/\lambda_+}(\beta_\epsilon\wedge\rho_\epsilon) + K\epsilon \\ &\leq \epsilon^{\alpha p}(1 + 2K\rho\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}|\log\epsilon| + K\epsilon^{1-\alpha p}),\end{aligned}$$

hold for all $\epsilon > 0$ small enough. Hence, for any small enough $\epsilon > 0$,

$$\mathbf{P}\{\beta_\epsilon < \rho_\epsilon\} = \mathbf{P}\left\{\sup_{t \le \beta_\epsilon \wedge \rho_\epsilon} |Y_{\epsilon,1}(t)| e^{-\lambda_+ t} \ge (1+\gamma)\epsilon^{\alpha p}\right\}$$
$$\le \mathbf{P}\left\{K\rho\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}|\log\epsilon| + K\epsilon^{1-\alpha p} \ge \gamma\right\} + \delta,$$

which implies the result. ■

The following is an important consequence of Lemma 42:

Corollary 44 *With τ_ϵ as in (2.25) it holds that*

$$\lim_{\epsilon\to 0} \mathbf{P}\{\tau_\epsilon^U < \tau_\epsilon\} = 0.$$

In particular, (2.27) holds.

From now on, we restrict our analysis to the high probability event $\{\tau_\epsilon^U \ge \tau_\epsilon\}$.

Let $\theta_\epsilon^+ = \epsilon^{-\alpha p} e^{-\lambda_+ \tau_\epsilon} Y_{\epsilon,1}(\tau_\epsilon)$. Then, (2.25) implies

$$\tau_\epsilon = -\frac{\alpha p}{\lambda_+}\log\epsilon + \frac{1}{\lambda_+}\log\frac{\delta}{|\theta_\epsilon^+|}, \tag{2.52}$$

and

$$Y_{\epsilon,1}(\tau_\epsilon) = \delta\,\mathrm{sgn}\,\theta_\epsilon^+.$$

Our analysis of these expressions will be based on the next formula which directly follows from Duhamel's principle:

$$\theta_\epsilon^+ = \mathrm{sgn}\,\eta_\epsilon^+ + \epsilon^{-\alpha p}\int_0^{\tau_\epsilon} e^{-\lambda_+ s} H_1(Y_\epsilon(s),\epsilon)ds + \epsilon^{1-\alpha p} N_\epsilon^+(\tau_\epsilon). \tag{2.53}$$

The main term in the r.h.s. of (2.53) is $\mathrm{sgn}\,\eta_\epsilon^+$. We need to estimate the other two terms. Lemma 36 implies that $\epsilon^{1-\alpha p}N_\epsilon^+(\tau_\epsilon)$ converges to 0 in probability as $\epsilon \to 0$. Let us now estimate the integral term. Relations (2.52) and (2.53) imply that $(\tau_\epsilon)_{\epsilon>0}$ is slowly growing, and we can use Lemma 42 to derive

$$\sup_{t\le\tau_\epsilon}|Y_{\epsilon,2}(t)| = O_{\mathbf{P}}(\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}). \tag{2.54}$$

We can now use Theorem 31 to conclude that

$$\epsilon^{-\alpha p}\sup_{t\le\tau_\epsilon}|H_1(Y_\epsilon(t),\epsilon)| = O_{\mathbf{P}}(\epsilon^{\alpha(1-p)\lambda_-/\lambda_+ - \alpha p} + \epsilon^{2-\alpha p}),$$

and (2.18) implies that the r.h.s. converges to 0. Therefore,

$$\epsilon^{-\alpha p} \int_0^{\tau_\epsilon} e^{-\lambda_+ s} H_1(Y_\epsilon(s), \epsilon) ds \xrightarrow{\mathbf{P}} 0.$$

The above analysis of equation (2.53) implies that if we define $\theta_0^+ = \operatorname{sgn} \eta_0^+$, then

$$\theta_\epsilon^+ \xrightarrow{Law} \theta_0^+, \tag{2.55}$$

which implies (2.28) due to (2.52). It remains to prove (2.29).

Duhamel's principle along with (2.52) yields

$$Y_{\epsilon,2}(\tau_\epsilon) = \left(\frac{|\theta_\epsilon^+|}{\delta}\right)^{\lambda_-/\lambda_+} \epsilon^{\alpha\lambda_-/\lambda_+} \eta_\epsilon^- + \int_0^{\tau_\epsilon} e^{-\lambda_-(\tau_\epsilon - s)} H_2(Y_\epsilon(s), \epsilon) ds + \epsilon N_\epsilon^-(\tau_\epsilon). \tag{2.56}$$

In order to study the convergence of $N_\epsilon^-(\tau_\epsilon)$ we first give a preliminary result.

Lemma 45

$$\sup_{t \le \tau_\epsilon} |Y_{\epsilon,1}(t) - \epsilon^{\alpha p} e^{\lambda_+ t} \operatorname{sgn} \eta_\epsilon^+| \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0.$$

Proof. The lemma follows from Duhamel's principle and Lemma 43. ∎

The following result is essentially Lemma 8.9 from [4]. It holds true in our setting since its proof is based only on the conclusion of Lemma 45.

Lemma 46 *As $\epsilon \to 0$,*

$$N_\epsilon^-(\tau_\epsilon) \xrightarrow{Law} N,$$

where N is the Gaussian random variable in (2.26).

We finish the proof of Lemma 33. Recall that the process M_ϵ was defined in (2.51) and introduce the stochastic processes

$$R_\epsilon(t) = \int_0^t e^{-\lambda_-(t-s)} \hat{H}_2(Y_\epsilon(s)) ds. \tag{2.57}$$

Note that (2.56) and (2.52) imply

$$\begin{aligned} Y_{\epsilon,2}(\tau_\epsilon) &= e^{-\lambda_- \tau_\epsilon} Y_{\epsilon,2}(0) + \int_0^{\tau_\epsilon} e^{-\lambda_-(\tau_\epsilon - s)} H_2(Y_\epsilon(s), \epsilon) ds + \epsilon N_\epsilon^-(\tau_\epsilon) \\ &= e^{-\lambda_- \tau_\epsilon} \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} \eta_\epsilon^- + \epsilon M_\epsilon(\tau_\epsilon) + R_\epsilon(\tau_\epsilon) \\ &= \eta_\epsilon^- \left(\frac{|\theta_\epsilon^+|}{\delta}\right)^{\lambda_-/\lambda_+} \epsilon^{\alpha\lambda_-/\lambda_+} + \epsilon M_\epsilon(\tau_\epsilon) + R_\epsilon(\tau_\epsilon). \end{aligned} \tag{2.58}$$

Relations (2.24) and (2.55) imply

$$\eta_\epsilon^- \left(\frac{|\theta_\epsilon^+|}{\delta}\right)^{\lambda_-/\lambda_+} \xrightarrow{Law} \left(\frac{|\eta_0^+|}{\delta}\right)^{\lambda_-/\lambda_+} y_2. \tag{2.59}$$

Lemma 46 and estimate (2.54) imply

$$M_\epsilon(\tau_\epsilon) \xrightarrow{Law} N, \quad \epsilon \to 0. \tag{2.60}$$

Equations (2.59) and (2.60) describe the behavior of first two terms in (2.58) and the proof of the lemma will be complete as soon as we show that

$$\epsilon^{-\beta} R_\epsilon(\tau_\epsilon) \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0. \tag{2.61}$$

We can write the following rough estimate based on (2.54) and Theorem 31:

$$\sup_{t \le \tau_\epsilon} |R_\epsilon(t)| = O_{\mathbf{P}}(\epsilon^{2\alpha(1-p)\lambda_-/\lambda_+}). \tag{2.62}$$

This is not sufficient for our purposes. We shall need a more detailed analysis instead. First, note that

$$\sup_{t \le \tau_\epsilon} |Y_{\epsilon,2}(t) - \epsilon M_\epsilon(t) - R_\epsilon(t)| e^{\lambda_- t} = \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} |\eta_\epsilon^-| = O_{\mathbf{P}}(\epsilon^{\alpha(1-p)\lambda_-/\lambda_+}).$$

Hence, for any $\gamma > 0$ there is a $K_\gamma > 0$ such that the event

$$D_\epsilon = \left\{ \sup_{t \le \tau_\epsilon} |Y_{\epsilon,2}(t) - \epsilon M_\epsilon(t) - R_\epsilon(t)| e^{\lambda_- t} < K_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} \right\}$$

has probability $\mathbf{P}(D_\epsilon) > 1 - \gamma$ for $\epsilon > 0$ small enough. Moreover, using Theorem 31 we see that for some constant $K_\beta > 0$,

$$|R_\epsilon(t)| \le K_\beta \int_0^t e^{-\lambda_-(t-s)} Y_{\epsilon,2}^2(s) ds.$$

Then, using the inequality $(a-b)^2 \le 2a^2 + 2b^2$ and defining $K_{\beta,\gamma} = K_\beta K_\gamma$, we see that on D_ϵ for each $t \le \tau_\epsilon$,

$$\begin{aligned} |R_\epsilon(t)| &\le K_\beta e^{-\lambda_- t} \int_0^t (e^{\lambda_- s} Y_{\epsilon,2}(s))^2 e^{-\lambda_- s} ds \\ &\le 2K_{\beta,\gamma} e^{-\lambda_- t} \int_0^t e^{-\lambda_- s} \epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} ds + 2K_\beta \int_0^t e^{-\lambda_-(t-s)} |\epsilon M_\epsilon(s) + R_\epsilon(s)|^2 ds \\ &\le 2\frac{K_{\beta,\gamma}}{\lambda_-} \epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} e^{-\lambda_- t} + 4\frac{K_\beta}{\lambda_-} \epsilon^2 M_{\epsilon,\infty}^2 + 4K_\beta e^{-\lambda_- t} \int_0^t e^{\lambda_- s} R_\epsilon(s)^2 ds, \end{aligned} \tag{2.63}$$

where $M_{\epsilon,\infty} = \sup_{t\le\tau_\epsilon} |M_\epsilon(t)|$, so that (according to Lemma 36) $M_{\epsilon,\infty}$ is slowly growing. Due to (2.62) we can find a constant $K'_\gamma > 0$ (independent of $\epsilon > 0$ and $t > 0$) so that the event

$$D'_\epsilon = D_\epsilon \cap \left\{ \sup_{t\le\tau_\epsilon} |R_\epsilon(t)| \le K'_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} \right\}$$

has probability $\mathbf{P}(D'_\epsilon) > 1-\gamma$ for all $\epsilon > 0$ small enough. Hence, multiplying both sides of (2.63) by $e^{\lambda_- t}$, we see that for some constant $C_\gamma > 0$ and all $t \le \tau_\epsilon$,

$$e^{\lambda_- t}|R_\epsilon(t)|\mathbf{1}_{\mathcal{D}'_\epsilon} \le \alpha(t) + C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} \int_0^t e^{\lambda_- s}|R_\epsilon(s)|\mathbf{1}_{\mathcal{D}'_\epsilon} ds,$$

where

$$\alpha(t) = C_\gamma \epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} + C_\gamma \epsilon^2 M^2_{\epsilon,\infty} e^{\lambda_- t}. \tag{2.64}$$

Using Gronwall's lemma and (2.64) we get

$$\begin{aligned}
\mathbf{1}_{\mathcal{D}'_\epsilon} e^{\lambda_- t}|R_\epsilon(t)| &\le \alpha(t) + C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} \int_0^t \alpha(s) e^{C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+}(t-s)} ds \\
&\le \alpha(t) + C_\gamma^2 \epsilon^{3\alpha(1-p)\lambda_-/\lambda_+} t e^{C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} t} \\
&\quad + \frac{C_\gamma^2}{\lambda_-} \epsilon^{2+\alpha(1-p)\lambda_-/\lambda_+} M^2_{\epsilon,\infty} t e^{\lambda_- t + C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+}}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbf{1}_{\mathcal{D}'_\epsilon}|R_\epsilon(t)| &\le C_\gamma \epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} e^{-\lambda_- t}(1 + C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} t e^{C_\gamma \epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} t}) \\
&\quad + C_\gamma \epsilon^2 M^2_{\epsilon,\infty}(1 + \frac{C_\gamma}{\lambda_-} \epsilon^{\alpha(1-p)\lambda_-/\lambda_+} t e^{C_\gamma \epsilon^{\alpha(1-p)\lambda_-/\lambda_+}}).
\end{aligned}$$

Using (2.52), we get that for any $q > 0$,

$$\begin{aligned}
\mathbf{1}_{\mathcal{D}'_\epsilon}|R_\epsilon(\tau_\epsilon)| &= O_{\mathbf{P}}\left(\epsilon^{2\alpha(1-p)\lambda_-/\lambda_+} e^{-\lambda_- \tau_\epsilon} + \epsilon^2 M^2_{\epsilon,\infty}\right) \\
&= O_{\mathbf{P}}\left(\epsilon^{\alpha\lambda_-/\lambda_+ + \alpha(1-p)\lambda_-/\lambda_+} + \epsilon^{2-q}\right),
\end{aligned}$$

so that (2.61) follows, and the proof is complete by choosing q small enough.

Chapter 3

Levinson Case

In this chapter we study Levinson case as presented in Section 1.3 of Chapter 1. We then apply the results obtained for this case to the 1-dimensional diffusion conditioned on rare events as explained in Section 1.4.1 of Chapter 1.

The chapter is organized as follows. In Section 3.2 we state the main theorem for the Levinson case, postponing its proof to Section 3.4. A approximation to the diffusion by the deterministic flow in finite time is presented in Section 3.3. This approximation is a key ingredient in all the arguments of Section 3.4. In Section 3.5 we state the result on the diffusion conditioned on a rare event and derive it from the main theorem and some auxiliary statements proven in Section 3.5.1.

3.1 Introduction

In this section we consider the dynamics in d dimensions. That is, we consider a C^2-smooth bounded vector field b in $\mathbb{R}^d$. The unperturbed dynamics is given by the deterministic flow $S = (S^t)_{t\in\mathbb{R}}$ generated by b.

The model has slight modifications from the classical exit problem. For this chapter, we introduce three components of perturbations of this deterministic flow. They all depend on a small parameter $\epsilon > 0$.

The first component is white noise perturbation generated by the matrix $\epsilon\sigma$, where $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a C^2-smooth bounded matrix valued function.

The second one is $\epsilon^{\alpha_1}\Psi_\epsilon$, where Ψ_ϵ is a deterministic Lipschitz vector field on $\mathbb{R}^d$ for each ϵ, converging uniformly to a limiting Lipschitz vector field Ψ_0, and α_1 is a positive scaling exponent. These conditions ensure that

the stochastic Itô equation

$$dX_\epsilon(t) = (b(X_\epsilon(t)) + \epsilon^{\alpha_1}\Psi_\epsilon(X_\epsilon(t)))\,dt + \epsilon\sigma(X_\epsilon(t))dW \tag{3.1}$$

w.r.t. a standard d-dimensional Wiener process W has a unique strong solution for any $\epsilon > 0$ and all initial conditions.

The last component of the perturbation is the initial condition satisfying

$$X_\epsilon(0) = x_0 + \epsilon^{\alpha_2}\xi_\epsilon, \quad \epsilon > 0. \tag{3.2}$$

Here $\alpha_2 > 0$, and $(\xi_\epsilon)_{\epsilon>0}$ is a family of random variables independent of W, such that for some random variable ξ_0, $\xi_\epsilon \to \xi_0$ as $\epsilon \to 0$ in distribution.

Let M be a C^2-smooth hypersurface in $\mathbb{R}^d$. If

$$\tau_\epsilon = \inf\{t \geq 0 : X_\epsilon(t) \in M\},$$

then on $\{\tau_\epsilon < \infty\}$ we have $X_\epsilon(\tau_\epsilon) \in M$. We are going to study the exit problem from M under the assumptions above. We use M instead of D, since M is assumed to be an hypersurface and we want to stick to the standard notation. In this setting, we state the main theorem in the next section.

3.2 Main result

In this section we state the main theorem and its hypothesis. Let us start with the assumptions on the joint geometry of the vector field b and the surface M. First we define

$$T = \inf\left\{t > 0 : S^t x_0 \in M\right\},$$

and assume that $0 < T < \infty$. Secondly, we denote $z = S^T x_0 \in M$ and assume that $b(z)$ does not belong to the tangent hyperplane $T_z M$. In other words, we assume that the positive orbit of x_0 intersects M and the crossing is transversal. The reader can check that this is equivalent to Condition 10 in Section 1.3 of Chapter 1.

In the case of $\xi_\epsilon \equiv 0$ and $\Psi \equiv 0$, Levinson's theorem states (see [46], [34, Chapter 2], and [35, Chapter 2]) that $X_\epsilon(\tau_\epsilon) \to z$ in probability as $\epsilon \to 0$. Levinson worked in the PDE context and showed how to obtain an expansion for the solution of the corresponding elliptic PDE depending on the small parameter ϵ. The main result of this note describes the limiting behavior of the correction $(\tau_\epsilon - T, X_\epsilon(\tau_\epsilon) - z)$ and extends [34, Theorem

2.3] to the situation with generic perturbation parameters ξ_0, Ψ, α_1, and α_2. This extension is essential since, as the analysis in [4] shows, in the sequential study of entrance-exit distributions for multiple domains one has to consider nontrivial scaling laws for the initial conditions; also, considering nontrivial deterministic perturbations will allow us to study rare events, see Section 3.5.

We need more notation. Due to the smoothness of b,

$$b(x) = b(y) + Db(y)(x-y) + Q_1(y, x-y), \quad x, y \in \mathbb{R}^d, \tag{3.3}$$

where

$$|Q_1(u,v)| \leq K|v|^2, \tag{3.4}$$

for some constant $K > 0$ and any $u, v \in \mathbb{R}^d$. We denote by $\Phi_x(t)$ the linearization of S along the orbit of x:

$$\frac{d}{dt}\Phi_x(t) = A(t)\Phi_x(t), \;\; \Phi_x(0) = I, \tag{3.5}$$

where $A(t) = Db(S^t x)$ and I is the identity matrix.

Finally, for any vector $v \in \mathbb{R}^d$, we define $\pi_b v \in \mathbb{R}$ and $\pi_M v \in T_z M$ by

$$v = \pi_b v \cdot b(z) + \pi_M v,$$

i.e., π_b is the (algebraic) projection onto span$(b(z))$ along $T_z M$ and π_M is the (geometric) projection onto $T_z M$ along span$(b(z))$.

Theorem 47 *Let $\alpha = \alpha_1 \wedge \alpha_2 \wedge 1$, and*

$$\begin{aligned}\phi_0(t) = \mathbf{1}_{\{\alpha_2=\alpha\}}\Phi_{x_0}(t)\xi_0 + \mathbf{1}_{\{\alpha_1=\alpha\}}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}\Psi_0(S^s x)ds \\ + \mathbf{1}_{\{1=\alpha\}}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}^{-1}(s)\sigma(S^s x_0)dW(s), \quad t > 0.\end{aligned} \tag{3.6}$$

Then, in the setting introduced above,

$$\epsilon^{-\alpha}(\tau_\epsilon - T, X_\epsilon(\tau_\epsilon) - z) \to (-\pi_b\phi_0(T), \pi_M\phi_0(T)). \tag{3.7}$$

in distribution. If additionally we require that $\xi_\epsilon \to \xi_0$ in probability or that $\alpha_2 > \alpha$, then the convergence in (3.7) *is also in probability.*

Remark 1 The conditions of Theorem 47 can be relaxed using the standard localization procedure. In fact, one needs to require uniform convergence of $\Psi_\epsilon \to \Psi_0$ and regularity properties of b and σ only in some neighborhood of the set $\{S^t x_0 : \; 0 \leq t \leq T(x_0)\}$.

Remark 2 In applications (see [4],[7]), the parameters α_1 and α_2 can be chosen so that the r.h.s. of (3.7) is nondegenerate.

Remark 3 In the case where $d = 1$, the hypersurface M is just a point. Therefore, π_M is identical zero and the only contentful information Theorem 47 provides is the asymptotics of the exit time.

3.3 A finite time approximation result

With high probability, at time T the process X_ϵ is close to z and the hitting time τ_ϵ is close to T. The idea behind the proof of Theorem 47 is that while the diffusion is close to z, the process may be approximated very well by motion with constant velocity $b(z)$.

In this section we prove the main ingredient to ensure this approximation.

Lemma 48 *Let X_ϵ be the solution of the SDE* (3.1) *with initial condition* (3.2). *Let*

$$\Theta_\epsilon(t) = \epsilon^{\alpha_2-\alpha}\Phi_{x_0}(t)\xi_\epsilon + \epsilon^{\alpha_1-\alpha}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}\Psi_0(S^s x_0)ds$$
$$+ \epsilon^{1-\alpha}\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}\sigma(S^s x_0)dW(s). \tag{3.8}$$

Then,

$$X_\epsilon(t) = S^t x_0 + \epsilon^\alpha \phi_\epsilon(t)$$

holds almost surely for every $t > 0$, where $\phi_\epsilon(t) = \Theta_\epsilon(t) + r_\epsilon(t)$, and r_ϵ converges to 0 uniformly over compact time intervals in probability.

If $\xi_\epsilon \to \xi_0$ in distribution, then for any $T > 0$, $\phi_\epsilon \to \phi_0$ in distribution in $C[0,T]$ equipped with uniform norm, where ϕ_0 is the stochastic process defined in (3.6).

If $\xi_\epsilon \to \xi_0$ in probability or $\alpha_2 > \alpha$, then the uniform convergence for ϕ_ϵ also holds in probability.

Remark 4 This lemma gives the first-order approximation for $X_\epsilon(t)$. Higher-order approximations in the spirit of [15] are also possible. They can be used to refine Theorem 47.

Proof. Let $\Delta_\epsilon^t = X_\epsilon(t) - S^t x_0$ and note that it satisfies the equation

$$d\Delta_\epsilon^t = \left(\left(b(X_\epsilon(t)) - b(S^t x_0)\right) + \epsilon^{\alpha_1}\Psi_\epsilon(X_\epsilon(t))\right)dt + \epsilon\sigma(X_\epsilon(t))dW(t),$$

with initial condition $\Delta_\epsilon^0 = \epsilon^{\alpha_2}\xi_\epsilon$. We want to study the properties of this equation. We start with the difference in b. Since b is a C^2 vector field, we may write

$$b(X_\epsilon(t)) - b(S^t x_0) = Db(S^t x_0)\Delta_\epsilon^t + Q_1(S^t x_0, \Delta_\epsilon^t). \tag{3.9}$$

Also, we can write

$$\Psi_\epsilon(X_\epsilon(t)) = \Psi_0(S^t x_0) + Q_2(S^t x_0, \Delta_\epsilon^t) + R_\epsilon(S^t x_0), \tag{3.10}$$

and

$$\sigma(X_\epsilon(t)) = \sigma(S^t x_0) + Q_3(S^t x_0, \Delta_\epsilon^t), \tag{3.11}$$

where

$$R_\epsilon(x) = \Psi_\epsilon(x) - \Psi_0(x) = o(1), \quad \epsilon \to 0,$$

uniformly in x; $Q_i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, $i = 2, 3$ satisfies

$$|Q_i(u, v)| \leq K|v|, \quad u, v \in \mathbb{R}^d. \tag{3.12}$$

We can assume that the constant $K > 0$ in (3.4) and (3.12) is the same for simplicity of notation.

Let $Q = Q_1 + \epsilon^{\alpha_1} Q_2 + \epsilon^{\alpha_1} R_\epsilon$. Combine (3.9), (3.10), and (3.11) to get

$$\begin{aligned} d\Delta_\epsilon^t = &\left(A(t)\Delta_\epsilon^t + \epsilon^{\alpha_1}\Psi_0(S^t x_0) + Q(S^t x_0, \Delta_\epsilon^t)\right) dt \\ &+ \epsilon\left(\sigma(S^t x_0) + Q_3(S^t x_0, \Delta_\epsilon^t)\right) dW(t), \qquad (3.13)\\ \Delta_\epsilon^0 = &\epsilon^{\alpha_2}\xi_\epsilon. \qquad (3.14) \end{aligned}$$

Hence, applying Duhamel's principle to (3.13) and using (3.8), we get

$$\begin{aligned} \Delta_\epsilon^t &= \epsilon^\alpha \Theta_\epsilon(t) + \Phi_{x_0}(t) \int_0^t \Phi_{x_0}(s)^{-1} Q(S^s x_0, \Delta_\epsilon^s) ds \\ &\quad + \epsilon \Phi_{x_0}(t) \int_0^t \Phi_{x_0}(s)^{-1} Q_3(S^s x_0, \Delta_\epsilon^s) dW(s) \\ &= \epsilon^\alpha \Theta_\epsilon(t) + \Theta'_\epsilon(t), \qquad (3.15) \end{aligned}$$

where Θ'_ϵ is defined by (3.15). A simple inspection of (3.8) shows that $(\Theta_\epsilon)_{\epsilon>0}$ converges in distribution in $C(0,T)$ to the process $\phi_0(t)$. This convergence is in probability if $\alpha_2 > \alpha$ or $\xi_\epsilon \to \xi_0$ in probability. Therefore, the lemma will follow with $\phi_\epsilon = \Theta_\epsilon + \epsilon^{-\alpha}\Theta'_\epsilon$ if we show that

$$\epsilon^{-\alpha} \sup_{t \leq T} |\Theta'_\epsilon(t)| \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0. \tag{3.16}$$

For any $\delta \in (1/2, 1)$, we introduce the stopping time

$$l_\epsilon(\delta) = \inf\left\{t > 0 : |\Delta_\epsilon^t| \geq \epsilon^{\alpha\delta}\right\}.$$

Now, $\Theta'_\epsilon = \Theta'_{\epsilon,1} + \epsilon\Theta'_{\epsilon,2}$, where

$$\Theta'_{\epsilon,1}(t) = \Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}Q(S^s x_0, \Delta_\epsilon^s)ds,$$

and

$$\Theta'_{\epsilon,2}(t) = \epsilon\Phi_{x_0}(t)\int_0^t \Phi_{x_0}(s)^{-1}Q_3(S^s x_0, \Delta_\epsilon^s)dW(s).$$

Bounds (3.4), and (3.12) imply

$$\sup_{t\leq T\wedge l_\epsilon(\delta)} |\Theta'_{\epsilon,1}(t)| = O(\epsilon^{2\alpha\delta} + \epsilon^{\alpha_1+\alpha\delta}) + o(\epsilon^{\alpha_1}) = o(\epsilon^\alpha). \tag{3.17}$$

Likewise, (3.12) for Q_3 and BDG inequality imply that for any $\kappa > 0$ there is a constant K_κ such that

$$\mathbf{P}\left\{\sup_{t\leq T\wedge l_\epsilon(\delta)} |\Theta'_{\epsilon,2}(t)| > K_\kappa\epsilon^{1+\alpha\delta}\right\} < \kappa \tag{3.18}$$

for all $\epsilon > 0$ small enough. Then, this together with (3.17) imply that

$$\epsilon^{-\alpha\delta}\sup_{t\leq T\wedge l_\epsilon(\delta)} |\Theta'_\epsilon(t)| \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0. \tag{3.19}$$

Then, if $l_\epsilon(\delta) < T$ we use (3.15) to get

$$\begin{aligned}1 &= \epsilon^{-\alpha\delta}\sup_{t\leq T\wedge l_\epsilon(\delta)} |\Delta_\epsilon^t| \\ &\leq \epsilon^{\alpha(1-\delta)}\sup_{t\leq T\wedge l_\epsilon(\delta)} |\Theta_\epsilon(t)| + \epsilon^{-\alpha\delta}\sup_{t\leq T\wedge l_\epsilon(\delta)} |\Theta'_\epsilon(t)|.\end{aligned}$$

The r.h.s. converges to 0 in probability due to (3.19) and the tightness of distributions of Θ_ϵ. Hence, $\mathbf{P}\{l_\epsilon(\delta) < T\} \to 0$ as $\epsilon \to 0$. Using T instead of $T \wedge l_\epsilon(\delta)$ in (3.17) and (3.18), we see that with the choice of $\delta > 1/2$, (3.16) follows and the proof is finished. ∎

3.4 Proof of Theorem 47

As we said before, the core idea of the proof is to approximate the behavior of the process X_ϵ with that of the deterministic flow in a small neighborhood of z. Let us start analyzing the process $X_\epsilon(t) - z$ for t close to T. Let us first estimate the deviation of the flow S from the motion with costant velocity $b(z)$. Let

$$r_\pm(t,x) = S^{\pm t}x - (x \pm tb(z)), \quad t > 0, \ x \in \mathbb{R}^d. \tag{3.20}$$

Lemma 49 *There are constants C_1 and C_2 so that for any $t > 0$ and $x \in \mathbb{R}^d$*

$$\sup_{s \le t} |r_\pm(s,x)| \le C_1 e^{C_2 t}(t|x-z| + t^2).$$

Proof. We prove the result for r_+. The analysis of r_- is similar since $S^{-t}x$ is the solution to the ODE

$$\frac{d}{dt} S^{-t}x = -b(S^{-t}x).$$

Let $L > 0$ be the Lipschitz constant of b. The proof follows from the inequalities:

$$\begin{aligned}
|r_+(t,x)| &\le \int_0^t |b(S^s x) - b(z)|\, ds \\
&\le L \int_0^t |S^s x - z|\, ds \\
&\le L \int_0^t |r_+(s,x)|\, ds + L \int_0^t |x + sb(z) - z|\, ds \\
&\le L \int_0^t |r_+(s,x)|\, ds + L \int_0^t |x - z| ds + L \int_0^t s|b(z)| ds \\
&\le L \int_0^t |r_+(s,x)|\, ds + Lt|x-z| + t^2 L|b(z)|/2.
\end{aligned}$$

The result follows as an application of Gronwall's lemma. ■

Lemma 50 *Let $\gamma \in (\alpha/2, \alpha)$. Then, there are two a.s.-continuous stochastic processes $\Gamma_{\epsilon,\pm}$ such that*

$$\sup_{t \in [0,\epsilon^\gamma]} |\Gamma_{\epsilon,\pm}(t)| \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0,$$

and almost surely for any $t \in [0, \epsilon^\gamma]$

$$X_\epsilon(T-t) = z - tb(z) + \epsilon^\alpha \left(\phi_\epsilon(T-t) + \Gamma_{\epsilon,-}(t)\right) \tag{3.21}$$

and

$$X_\epsilon(T+t) = z + tb(z) + \epsilon^\alpha \left(\Phi_z(t)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(t)\right). \tag{3.22}$$

Proof. Due to Lemma 48, the flow property, and (3.20) we have

$$\begin{aligned} X_\epsilon(T-t) &= S^{T-t}x_0 + \epsilon^\alpha \phi_\epsilon(T-t) \\ &= S^{-t}z + \epsilon^\alpha \phi_\epsilon(T-t) \\ &= z - tb(z) + r_-(t,z) + \epsilon^\alpha \phi_\epsilon(T-t). \end{aligned}$$

The first estimate with $\Gamma_{\epsilon,-}(t) = \epsilon^{-\alpha} r_-(t,z)$ follows from Lemma 49 for $x = z$.

Due to Strong Markov property and Lemma 47 the process $\tilde{X}_\epsilon(t) = X_\epsilon(t+T)$ is a solution of the initial value problem

$$\begin{aligned} d\tilde{X}_\epsilon(t) &= (b(\tilde{X}_\epsilon(t)) + \epsilon^{\alpha_1}\Psi_\epsilon(\tilde{X}_\epsilon(t)))dt + \epsilon\sigma(\tilde{X}_\epsilon(t))d\tilde{W}, \\ \tilde{X}_\epsilon(0) &= X_\epsilon(T) = z + \epsilon^\alpha \phi_\epsilon(T), \end{aligned}$$

with respect to the Brownian Motion $\tilde{W}(t) = W(t+T) - W(T)$. So, again, applying Lemma 47 to this shifted equation, we obtain $\tilde{X}_\epsilon(t) = S^t z + \epsilon^\alpha \hat{\phi}_\epsilon(t)$, where, for $t > 0$

$$\hat{\phi}_\epsilon(t) = \Phi_z(t)\phi_\epsilon(T) + \theta_\epsilon(t),$$

and

$$\theta_\epsilon(t) = \epsilon^{1-\alpha}\Phi_z(t)\int_0^t \Phi_z(s)^{-1}\sigma(S^s z)d\tilde{W}(s) + \epsilon^{\alpha_1-\alpha}\Phi_z(t)\int_0^t \Phi_z(s)^{-1}\Psi_0(S^s z)ds + \tilde{r}_\epsilon(t),$$

where $\tilde{r}_\epsilon$ converges to 0 uniformly over compact time intervals in probability. Then due to (3.20),

$$\begin{aligned} \tilde{X}_\epsilon(t) &= S^t z + \epsilon^\alpha(\Phi_z(t)\phi_\epsilon(t) + \theta_\epsilon(t)) \\ &= z + tb(z) + r_+(t,z) + \epsilon^\alpha(\Phi_z(t)\phi_\epsilon(t) + \theta_\epsilon(t)). \end{aligned}$$

Hence, with $\Gamma_{\epsilon,+}(t) = \theta_\epsilon(t) + \epsilon^{-\alpha} r_+(t,z)$ the result is a consequence of Lemma 49. ■

Let us now parametrize, locally around z, the hypersurface M as a graph of a C^2-function F over $T_z M$, i.e., $y \mapsto z + y + F(y)\cdot b(z)$ gives a C^2-parametrization of a neighborhood of z in M by a neighborhood of 0 in

T_zM. Moreover, $DF(0) = 0$ so that $|F(y)| = O(|y|^2)$, $y \to 0$. With this definition, it is clear that, for $w \in \mathbb{R}^d$ with $w - z$ small enough, $w \in M$ if and only if $\pi_b(w - z) = F(\pi_M(w - z))$.

Let us define

$$\begin{aligned}
\Omega_{1,\epsilon} &= \left\{\tau_\epsilon = \inf\{t \geq 0 : \pi_b\left(X_\epsilon(t) - z\right) = F\left(\pi_M\left(X_\epsilon(t) - z\right)\right)\}\right\}, \\
\Omega_{2,\epsilon} &= \left\{|\tau_\epsilon - T| \leq \epsilon^\gamma\right\}, \\
\Omega_\epsilon &= \Omega_{1,\epsilon} \cap \Omega_{2,\epsilon}.
\end{aligned}$$

Lemma 51 $\mathbf{P}(\Omega_\epsilon) \to 1$ *as* $\epsilon \to 0$.

Proof. The definition of F and Lemma 47 imply that as $\epsilon \to 0$, $\mathbf{P}(\Omega_{1,\epsilon}) \to 1$.

We use (3.22) to conclude that

$$\pi_b\left(X_\epsilon(T + \epsilon^\gamma) - z\right) = \epsilon^\gamma\left(1 + \epsilon^{\alpha-\gamma}\pi_b\left(\Phi_z(\epsilon^\gamma)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\epsilon^\gamma)\right)\right),$$

and

$$F\left(\pi_M\left(X_\epsilon(T + \epsilon^\gamma) - z\right)\right) = F\left(\epsilon^\alpha\pi_M\left(\Phi_z(\epsilon^\gamma)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\epsilon^\gamma)\right)\right).$$

Since $|F(x)| = O(|x|^2)$, these estimates imply that

$$\begin{aligned}
&\limsup_{\epsilon\to 0}\mathbf{P}\left(\{\tau_\epsilon > T + \epsilon^\gamma\} \cap \Omega_{1,\epsilon}\right) \\
&\qquad \leq \limsup_{\epsilon\to 0}\mathbf{P}\left\{\pi_b\left(X_\epsilon(T + \epsilon^\gamma) - z\right) \leq F\left(\pi_M\left(X_\epsilon(T + \epsilon^\gamma) - z\right)\right)\right\} = 0.
\end{aligned}$$

It remains to prove

$$\lim_{\epsilon\to 0}\mathbf{P}\left\{\tau_\epsilon < T - \epsilon^\gamma\right\} = 0. \tag{3.23}$$

Let us denote the Hausdorff distance between sets by $d(\cdot,\cdot)$. Then an obvious estimate

$$d(\{S^t x_0 : 0 \leq t \leq T - \delta\}, M) \geq c\delta$$

holds true for some $c > 0$ and all sufficiently small $\delta > 0$. Now (3.23) follows from Lemma 47, and the proof is complete ■

Lemma 52 *Define* $\tau'_\epsilon = \tau_\epsilon - T$. *Then,*

$$\epsilon^{-\alpha}\tau'_\epsilon + \pi_b\phi_\epsilon(T) \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0.$$

Proof. Let us define $A_\epsilon = \{0 \le \tau'_\epsilon \le \epsilon^\gamma\} \cap \Omega_{1,\epsilon}$ and $B_\epsilon = \{-\epsilon^\gamma \le \tau'_\epsilon < 0\} \cap \Omega_{1,\epsilon}$, so that $\Omega_\epsilon = A_\epsilon \cup B_\epsilon$. We can use (3.22) and the definition of $\Omega_{1,\epsilon}$ to get

$$\mathbf{1}_{A_\epsilon}\tau'_\epsilon + \mathbf{1}_{A_\epsilon}\epsilon^\alpha \pi_b\left(\Phi_z(\tau'_\epsilon)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\tau'_\epsilon)\right) = \mathbf{1}_{A_\epsilon} F\left(\epsilon^\alpha \pi_M\left(\Phi_z(\tau'_\epsilon)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\tau'_\epsilon)\right)\right).$$

This implies

$$\begin{aligned}
\mathbf{1}_{A_\epsilon}\epsilon^{-\alpha}\tau'_\epsilon &= \epsilon^{-\alpha}\mathbf{1}_{A_\epsilon} F\left(\epsilon^\alpha \pi_M\left(\Phi_z(\tau'_\epsilon)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\tau'_\epsilon)\right)\right) \\
&\quad - \mathbf{1}_{A_\epsilon}\pi_b\left(\Phi_z(\tau'_\epsilon)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\tau'_\epsilon)\right) \\
&= -\mathbf{1}_{A_\epsilon}\pi_b\left(\Phi_z(\tau'_\epsilon)\phi_\epsilon(T)\right) + r_{\epsilon,1} \\
&= -\mathbf{1}_{A_\epsilon}\pi_b\phi_\epsilon(T) + \mathbf{1}_{A_\epsilon}\pi_b\left((I - \Phi_z(\tau'_\epsilon))\phi_\epsilon(T)\right) + r_{\epsilon,1}, \qquad (3.24)
\end{aligned}$$

where $r_{\epsilon,1}$ is a random variable that converges to 0 in probability as $\epsilon \to 0$.

Likewise, since $\tau_\epsilon = T - (-\tau'_\epsilon)$ and $\mathbf{1}_{B_\epsilon}\tau'_\epsilon \le 0$, we can use (3.21) and the definition of $\Omega_{1,\epsilon}$ to see that

$$\mathbf{1}_{B_\epsilon}\tau'_\epsilon + \mathbf{1}_{B_\epsilon}\epsilon^\alpha\pi_b\left(\phi_\epsilon(T+\tau'_\epsilon) + \Gamma_{\epsilon,-}(-\tau'_\epsilon)\right) = \mathbf{1}_{B_\epsilon} F\left(\epsilon^\alpha\left(\phi_\epsilon(T+\tau'_\epsilon) + \Gamma_{\epsilon,-}(-\tau'_\epsilon)\right)\right).$$

Hence, proceeding as before, we see that

$$\begin{aligned}
\mathbf{1}_{B_\epsilon}\epsilon^{-\alpha}\tau'_\epsilon &= -\mathbf{1}_{B_\epsilon}\pi_b\phi_\epsilon(T+\tau'_\epsilon) + r_{\epsilon,2} \\
&= -\mathbf{1}_{B_\epsilon}\pi_b\phi_\epsilon(T) + \mathbf{1}_{B_\epsilon}\pi_b\left(\phi_\epsilon(T) - \phi_\epsilon(T+\tau'_\epsilon)\right) + r_{\epsilon,2}
\end{aligned}$$

for some random variable $r_{\epsilon,2}$ such that $r_{\epsilon,2} \to 0$ in probability as $\epsilon \to 0$. Adding this identity and (3.24), we see that on Ω_ϵ

$$\epsilon^{-\alpha}\tau'_\epsilon = -\pi_b\phi_\epsilon(T) + \mathbf{1}_{A_\epsilon}\pi_b\left((I - \Phi_z(\tau'_\epsilon))\phi_\epsilon(T)\right) + \mathbf{1}_{B_\epsilon}\pi_b\left(\phi_\epsilon(T) - \phi_\epsilon(T+\tau'_\epsilon)\right) + r_{\epsilon,1} + r_{\epsilon,2}.$$

Due to Lemma 51, to finish the proof it is sufficient to notice that as $\epsilon \to 0$

$$\sup_{0 \le t \le \epsilon^\gamma} |(I - \Phi_z(t))\phi_\epsilon(T)| \xrightarrow{\mathbf{P}} 0, \qquad (3.25)$$

and

$$\sup_{0 \le t \le \epsilon^\gamma} |\phi_\epsilon(T) - \phi_\epsilon(T+t)| \xrightarrow{\mathbf{P}} 0. \qquad (3.26)$$

■

Lemma 52 takes care of the time component in Theorem 47. We shall consider the spatial component now.

Let A_ϵ and B_ϵ be as in the proof of Lemma 52. Then, (3.22) implies

$$\begin{aligned}
&\mathbf{1}_{A_\epsilon}\left(X_\epsilon(\tau_\epsilon) - z\right)\epsilon^{-\alpha} = \mathbf{1}_{A_\epsilon}\epsilon^{-\alpha}\tau'_\epsilon b(z) + \mathbf{1}_{A_\epsilon}\left(\Phi_z(\tau'_\epsilon)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\tau'_\epsilon)\right) \\
&\quad = \mathbf{1}_{A_\epsilon}\left(\epsilon^{-\alpha}\tau'_\epsilon b(z) + \phi_\epsilon(T)\right) + \mathbf{1}_{A_\epsilon}\left[(\Phi_z(\tau'_\epsilon) - I)\phi_\epsilon(T) + \Gamma_{\epsilon,+}(\tau'_\epsilon)\right] \qquad (3.27)
\end{aligned}$$

Likewise, from (3.21) we get that

$$\begin{aligned}\mathbf{1}_{B_\epsilon}\left(X_\epsilon(\tau_\epsilon)-z\right)\epsilon^{-\alpha} &= \mathbf{1}_{B_\epsilon}\epsilon^{-\alpha}\tau'_\epsilon b(z)+\mathbf{1}_{B_\epsilon}\left(\phi_\epsilon(T+\tau'_\epsilon)+\Gamma_{\epsilon,-}(-\tau'_\epsilon)\right)\\ &= \mathbf{1}_{B_\epsilon}\left(\epsilon^{-\alpha}\tau'_\epsilon b(z)+\phi_\epsilon(T)\right)+\mathbf{1}_{B_\epsilon}\left[\left(\phi_\epsilon(T+\tau'_\epsilon)-\phi_\epsilon(T)\right)+\Gamma_{\epsilon,-}(-\tau'_\epsilon)\right].\end{aligned} \tag{3.28}$$

Adding (3.27) and (3.28) and proceding as in the proof of Lemma 52 we see that

$$\left(X_\epsilon(\tau_\epsilon)-z\right)\epsilon^{-\alpha}-\pi_M\phi_\epsilon(T)=\left(\epsilon^{-\alpha}\tau'_\epsilon+\pi_b\phi_\epsilon(T)\right)b(z)+\rho_\epsilon,$$

where, due to (3.25), (3.26) and Lemma 50, $\rho_\epsilon\to 0$ in probability as $\epsilon\to 0$. From this expression and Lemma 52 we get that

$$\left(X_\epsilon(\tau_\epsilon)-z\right)\epsilon^{-\alpha}-\pi_M\phi_\epsilon(T)\xrightarrow{\mathbf{P}}0,\quad \epsilon\to 0.$$

Then, using this and the convergence in Lemma 52

$$\epsilon^{-\alpha}(\tau_\epsilon-T,X_\epsilon(\tau_\epsilon)-z)=R_\epsilon+G(\phi_\epsilon(T)),$$

where R_ϵ is a random variable such that $R_\epsilon\to 0$ in probability as $\epsilon\to 0$. G is the continuous function $x\mapsto(-\pi_b x,\pi_M x)$. Hence, Theorem 47 follows from the convergence in Lemma 47.

3.5 Conditioned diffusions in 1 dimension

In this section we apply Theorem 47 to the analysis of the exit time of conditioned diffusions in 1-dimensional situation for the large deviation case.

Suppose, for each $\epsilon>0$, X_ϵ is a weak solution of the following SDE:

$$\begin{aligned}dX_\epsilon(t)&=b(X_\epsilon(t))dt+\epsilon\sigma(X_\epsilon(t))dW(t),\\ X_\epsilon(0)&=x_0,\end{aligned}$$

where b and σ are C^1 functions on $\mathbb{R}$, such that $b(x)<0$ and $\sigma(x)\neq 0$ for all x in an interval $[a_1,a_2]$ containing x_0. We introduce

$$\tau_\epsilon=\inf\{t\geq 0:\ X_\epsilon(t)=a_1\text{ or }a_2\}$$

and $B_\epsilon=\{X_\epsilon(\tau_\epsilon)=a_2\}$. Since $b<0$, B_ϵ is a rare event since $\lim_{\epsilon\to 0}\mathbf{P}(B_\epsilon)=0$. More precise estimates on the asymptotic behavior of $\mathbf{P}(B_\epsilon)$ can be obtained in terms of large deviations. However, here we study the diffusion X_ϵ conditioned on the rare event B_ϵ.

Let $T(x_0)$ denote the time it takes for the solution of $\dot{x} = -b(x)$ starting at x_0 to reach a_2. Given that $b < 0$ on the hole interval $[a_1, a_2]$, a simple calculation shows that

$$T(x_0) = -\int_{x_0}^{a_2} \frac{1}{b(x)} dx.$$

Theorem 53 *Conditioned on B_ϵ, the distribution of $\epsilon^{-1}(\tau_\epsilon - T(x_0))$ converges weakly to a centered Gaussian distribution with variance*

$$-\int_{x_0}^{a_2} \frac{\sigma^2(y)}{b^3(y)} dy.$$

To prove this theorem, we will need two auxiliary statements. Their proofs are given in Section 3.5.1.

Lemma 54 *Conditioned on B_ϵ, the process X_ϵ is a diffusion with the same diffusion coefficient as the unconditioned process, and with the drift coefficient given by*

$$b_\epsilon(x) = b(x) + \epsilon^2 \sigma^2(x) \frac{h_\epsilon(x)}{\int_{a_1}^{x} h_\epsilon(y) dy},$$

where

$$h_\epsilon(x) = \exp\left\{ -\frac{2}{\epsilon^2} \int_{a_1}^{x} \frac{b(y)}{\sigma^2(y)} dy \right\}. \tag{3.29}$$

Further analysis requires understanding the limiting behavior of b_ϵ. This is the purpose of the next lemma:

Lemma 55 *There is $\delta > 0$ such that*

$$\limsup_{\epsilon \to 0} \epsilon^{-2} \left(\sup_{x \in [x_0 - \delta, a_2 + \delta]} |b_\epsilon(x) + b(x)| \right) < \infty.$$

Remark 5 Although we need the condition that $b(x) < 0$ for all $x \in [a_1, a_2]$ for Theorem 53 to hold, Lemmas 54 and 55 hold independently of the sign properties of b.

Proof of Theorem 53. Let us fix $\beta \in (1, 2)$. Lemmas 54 and 55 imply that X_ϵ conditioned on B_ϵ, up to τ_ϵ satisfies an SDE of the form

$$dX_\epsilon(t) = \Big(-b(X_\epsilon(t)) + \epsilon^\beta \Psi_{\epsilon,\beta}(X_\epsilon(t)) \Big) dt + \epsilon \sigma(X_\epsilon(t)) d\tilde{W}(t),$$

for some Brownian Motion $\tilde{W}$ and with $\Psi_{\epsilon,\beta} \to 0$ uniformly as $\epsilon \to 0$. We can assume that after time τ_ϵ, this process still follows the same equation at least up to the time it hits $x_0 - \delta$ or $a_1 + \delta$.

So, having the dynamics from $\dot{x} = -b(x)$ as the underperturbed dynamics, we can apply Theorem 47 (taking into account Remark 1) to see that

$$\epsilon^{-1}(\tau_\epsilon - T(x_0)) \xrightarrow{\mathbf{P}} -\frac{1}{b(a_2)}\Phi_{x_0}(T(x_0)) \int_0^{T(x_0)} \Phi_{x_0}^{-1}(s)\sigma(S^s x_0)d\tilde{W}(s), \quad \epsilon \to 0, \tag{3.30}$$

where $S^t x_0$ is the flow generated by the vector field $-b$, the time $T(x_0)$ solves $S^{T(x_0)}x_0 = a_2$, and Φ_{x_0} is the linearization of S near the orbit of x_0. The limit is clearly a centered Gaussian random variable. To compute its variance we must first solve

$$\frac{d}{dt}\Phi_{x_0}(t) = -b'(S^t x_0)\Phi_{x_0}(t), \quad \Phi_{x_0}(0) = 1.$$

The solution to this linear ODE is

$$\Phi_{x_0}(t) = \exp\left\{-\int_0^t b'(S^s x_0)ds\right\},$$

so that after the change of variables $u = S^s x_0$ in the integral, we get

$$\Phi_{x_0}(t) = \frac{b(S^t x_0)}{b(x_0)}.$$

Using this expression and Itô isometry for the limiting random variable in (3.30), we get that the variance of such random variable is

$$\int_0^{T(x_0)} \frac{\sigma^2(S^t x_0)}{b^2(S^t x_0)}dt.$$

We can now use the change of variable $u = S^s x_0$ to get the expression in Theorem 53. ■

3.5.1 Proof of Lemmas 54 and 55

Proof of Lemma 54. Let us find the generator of the conditioned diffusion. To that end we denote the generator of the original diffusion by L_ϵ:

$$L_\epsilon f(x) = b(x)f'(x) + \frac{\epsilon^2}{2}\sigma^2(x)f''(x) = \lim_{t\to 0}\frac{\mathbf{E}_x f(X_\epsilon) - f(x)}{t}, \tag{3.31}$$

where f is any bounded C^2-function with bounded first two derivatives and $\mathbf{E}_x$ denotes expectation with respect to the measure $\mathbf{P}_x$, the element of the Markov family describing the Markov process emitted from point x.

Let us denote $u_\epsilon(x) = \mathbf{P}_x(B_\epsilon)$. This function solves the following boundary-value problem for the backward Kolmogorov equation:

$$L_\epsilon u_\epsilon(x) = 0, \quad u_\epsilon(a_1) = 0, \quad u_\epsilon(a_2) = 1.$$

Using (3.31), it is easy to check that a unique solution is given by

$$u_\epsilon(x) = \frac{\int_{a_1}^{x} h_\epsilon(y)dy}{\int_{a_1}^{a_2} h_\epsilon(y)dy},$$

where h_ϵ is defined in (3.29).

Now we can compute the generator $\bar{L}_\epsilon$ of the conditioned flow. For any smooth and bounded function $f \in C^2$ with bounded first two derivatives, we can write

$$\begin{aligned}
\mathbf{E}_x[f(X_\epsilon)|B_\epsilon] &= u_\epsilon^{-1}(x)\mathbf{E}_x f(X_\epsilon(t))\mathbf{1}_{B_\epsilon} \\
&= u_\epsilon^{-1}(x)\mathbf{E}_x f(X_\epsilon(t))\mathbf{1}_{B_\epsilon}\mathbf{1}_{\{\tau_\epsilon \geq t\}} + R_\epsilon \\
&= u_\epsilon^{-1}(x)\mathbf{E}_x\mathbf{E}_x[f(X_\epsilon(t))\mathbf{1}_{B_\epsilon}\mathbf{1}_{\{\tau_\epsilon \geq t\}}|\mathcal{F}_t] + R_\epsilon \\
&= u_\epsilon^{-1}(x)\mathbf{E}_x f(X_\epsilon(t))\mathbf{P}_{X_\epsilon(t)}(B_\epsilon) + R_\epsilon \\
&= u_\epsilon^{-1}(x)\mathbf{E}_x f(X_\epsilon(t))u_\epsilon(X_\epsilon(t)) + R_\epsilon,
\end{aligned}$$

where

$$|R_\epsilon| = u_\epsilon^{-1}(x)|\mathbf{E}_x f(X_\epsilon)\mathbf{1}_{B_\epsilon}\mathbf{1}_{\{\tau_\epsilon < t\}}| \leq C(x)\mathbf{P}\{\tau_\epsilon < t\} = o(t)$$

for some $C(x) > 0$. Therefore, we obtain

$$\begin{aligned}
\bar{L}_\epsilon f(x) &= \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_\epsilon(t))|B_\epsilon] - f(x)}{t} \\
&= \lim_{t\to 0} \frac{u_\epsilon^{-1}(x)\mathbf{E}_x f(X_\epsilon(t))u_\epsilon(X_\epsilon(t)) - f(x)}{t} \\
&= \frac{1}{u_\epsilon(x)} \lim_{t\to 0} \frac{\mathbf{E}_x f(X_\epsilon(t))u_\epsilon(X_\epsilon(t)) - f(x)u_\epsilon(x)}{t} \\
&= \frac{1}{u_\epsilon(x)} L_\epsilon(fu_\epsilon)(x) \\
&= \left(b(x) + \epsilon^2\sigma^2(x)\frac{u_\epsilon'(x)}{u_\epsilon(x)}\right) f'(x) + \epsilon^2\frac{\sigma^2(x)}{2}f''(x). \\
&= \left(b(x) + \epsilon^2\sigma^2(x)\frac{h_\epsilon(x)}{\int_{a_1}^{x} h_\epsilon(y)dy}\right) f'(x) + \epsilon^2\frac{\sigma^2(x)}{2}f''(x),
\end{aligned}$$

completing the proof. ■

Proof of Lemma 55. The proof is a variation of Laplace's method. Let

$$\Phi(x) = 2\int_{a_1}^{x} \frac{b(y)}{\sigma^2(y)} dy, \quad x \geq a_1, \tag{3.32}$$

so that $h_\epsilon(x) = e^{-\Phi(x)/\epsilon^2}$. We take any $\beta \in (1,2)$ and break the integral of h_ϵ in two parts:

$$\int_{a_1}^{x} e^{-\Phi(y)/\epsilon^2} dy = I_{\epsilon,1}(x) + I_{\epsilon,2}(x),$$

where

$$I_{\epsilon,1}(x) = \int_{a_1}^{x-\epsilon^\beta} e^{-\Phi(y)/\epsilon^2} dy, \tag{3.33}$$

and

$$I_{\epsilon,2}(x) = \int_{x-\epsilon^\beta}^{x} e^{-\Phi(y)/\epsilon^2} dy. \tag{3.34}$$

The idea is to prove that $I_{\epsilon,1}$ is exponentially smaller than $I_{\epsilon,2}$ and then estimate $I_{\epsilon,2}$.

We start with some preliminaries for the function Φ. Since both b and σ are C^1 and $\sigma \neq 0$ in $[a_1, a_2]$ we conclude that Φ is a C^2 function so that we can find a function $R : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and a number $\delta_0 > 0$ such that for every $x, y \in [a_1, a_2 + \delta_0]$, we have the expansion

$$\Phi(y) = \Phi(x) + \Phi'(x)(y - x) + R(x, y - x), \tag{3.35}$$

and

$$|R(x,v)| \leq K_1 |v|^2, \quad x \in [a_1, a_2 + \delta_0], v \in \mathbb{R}, \tag{3.36}$$

for some $K_1 > 0$.

To estimate $I_{\epsilon,1}$, we introduce

$$J_{\epsilon,1}(x) = \frac{e^{\Phi(x)/\epsilon^2}}{\epsilon^2 \sigma^2(x)} I_{\epsilon,1}(x), \quad x \in [a_1, a_2 + \delta_0].$$

Since Φ is decreasing, we have that for some constant $K_2 > 0$ independent of $x \in [a_1, a_2 + \delta_0]$,

$$J_{\epsilon,1}(x) \leq \frac{K_2}{\epsilon^2} e^{(\Phi(x) - \Phi(x-\epsilon^\beta))/\epsilon^2}. \tag{3.37}$$

Since $\beta < 2$ and Φ' is negative and bounded away from zero, we conclude that there is $\alpha(\epsilon)$ such that $\alpha(\epsilon) = o(\epsilon^2)$ as $\epsilon \to 0$ and

$$\sup_{x \in [a_1, a_2 + \delta_0]} J_{\epsilon,1}(x) \leq \alpha(\epsilon). \tag{3.38}$$

We now estimate $I_{\epsilon,2}$. Using expansion (3.35) and the change of variables $u = -\Phi(x)(y-x)/\epsilon^2$, we get

$$\begin{aligned}
I_{\epsilon,2}(x) &= e^{-\Phi(x)/\epsilon^2} \int_{x-\epsilon^\beta}^{x} e^{-\Phi'(x)(y-x)/\epsilon^2 - R(x,y-x)/\epsilon^2} dy \\
&= -\frac{\epsilon^2}{\Phi'(x)} e^{-\Phi(x)/\epsilon^2} \int_{\Phi'(x)/\epsilon^{2-\beta}}^{0} e^{u-R(x,-\epsilon^2 u/\Phi'(x))/\epsilon^2} du \\
&= -\frac{\epsilon^2\sigma^2(x)}{2b(x)} e^{-\Phi(x)/\epsilon^2} J_{\epsilon,2}(x), \qquad (3.39)
\end{aligned}$$

where we use (3.32) to compute the derivative of Φ, and we define $J_{\epsilon,2}$ by (3.39). Hence, combining (3.37) with the definition of b_ϵ and (3.39), we get

$$b_\epsilon(x) = b(x) + \frac{1}{J_{\epsilon,1}(x) - \frac{1}{2b(x)} J_{\epsilon,2}(x)}.$$

Due to (3.38), the proof will be complete once we prove that for sufficiently small $\delta > 0$,

$$\limsup_{\epsilon\to 0} \epsilon^{-2} \left(\sup_{x\in[x_0-\delta, a_2+\delta]} |J_{\epsilon,2}(x) - 1| \right) < \infty.$$

Note that for any $\delta \in (0, x_0 - a_1)$, some constant $K_3 = K_3(\delta) > 0$ and all $x \in [x_0 - \delta, a_2 + \delta]$,

$$\begin{aligned}
|J_{\epsilon,2}(x) - 1| &= \Big| \int_{\Phi'(x)/\epsilon^{2-\beta}}^{0} e^u (1 - e^{-R(x,-\epsilon^2 u/\Phi'(x))/\epsilon^2}) du \\
&\quad + \int_{-\infty}^{\Phi'(x)/\epsilon^{2-\beta}} e^u du \Big| \\
&\le \int_{\Phi'(x)/\epsilon^{2-\beta}}^{0} e^u |1 - e^{-R(x,-\epsilon^2 u/\Phi'(x))/\epsilon^2}| du + e^{-K_3/\epsilon^{2-\beta}}. \qquad (3.40)
\end{aligned}$$

Using (3.36) we see that for some constant $K_4 > 0$ independent of $x \in [x_0 - \delta, a_2 + \delta]$ and $u \in \mathbb{R}$,

$$|R(x, -\epsilon^2 u/\Phi'(x))|/\epsilon^2 \le K_4 \epsilon^2 u^2.$$

In particular,

$$\sup_{x\in[x_0-\delta,a_2+\delta]} \sup_{u\in[\Phi'(x)/\epsilon^{2-\beta},0]} |R(x, -\epsilon^2 u/\Phi'(x))|/\epsilon^2 \le K_4 \epsilon^{2(\beta-1)}.$$

Since $\beta > 1$, the r.h.s. converges to 0 and we can apply a basic Taylor estimate which implies that for all $\epsilon > 0$ small enough,

$$\sup_{x\in[x_0-\delta,a_2+\delta]} \sup_{u\in[\Phi'(x)/\epsilon^{2-\beta},0]} |1 - e^{-R(x,-\epsilon^2 u/\Phi'(x))/\epsilon^2}| \leq K_5\epsilon^2 u^2,$$

for some $K_5 > 0$. Using this fact in the integral of (3.40), we can find a constant $K_6 = K_6(\delta) > 0$ such that

$$\sup_{x\in[x_0-\delta,a_2+\delta]} |J_{\epsilon,2}(x) - 1| \leq K_6\epsilon^2 + e^{-K_3/\epsilon^{2-\beta}},$$

which finishes the proof. ■

Chapter 4

Conclusion

This chapter is devoted to give further discussion of the topics covered in this text. In Section 4.1 we made some comments related to the application of normal forms. In Section 4.2 we comment about the exit problem in the case where the deterministic flow has a unique saddle point. In Section 4.3 we present an open problem related to scaling limit and show a possible relation with Chapter 3.

4.1 Normal Forms

In this thesis, a transformation (normal form transformation) is used to conjugate the original equation for X_ϵ into a non-linear perturbation of the linearized equation. This is done so we can avoid an approximation step between our original equation and its linearization. There is evidence that a similar methodology has been in the mind of the researchers since the publication of [22]. A concrete conjugation of the original equation into the linearized system was used in [4]. Although this result was successful, it required certain assumptions that are removed in this work (in the 2-dimensional setting) by conjugating to a non-linear system instead. As far as the author knows, it is the first time a program of this nature has been successful.

Inspired by [13], in [30] a normal form transformation was applied to an epidemiological model. Contrastingly to our case, this is a specific equation and not an abstract setting. In [3] the normal form theory is presented for stochastic differential equations in the abstract setting. Although the transformation is presented, no estimates are computed as in this work.

Further the development of normal form theory in [3] is not complete. For example, it excludes the one-resonant case.

Although normal form theory has proved to be a powerful tool in dynamical systems, in probability is still not clear how powerful the theory really is. In this text we use the very explicit shape of the nonlinearity in the normal form to obtain specific estimates that successfully lead to a complete solution of the problem in Chapter 2. As far as the author knows, this work is the first time in which normal form theory is applied in an abstract setting and is used to obtain tight estimates that lead to a solution of a probabilistic problem. The approach presented has some further generalizations in which the application of normal forms may be useful. We give a brief presentation about the possible complications that may be found.

4.2 Escape from a Saddle: further generalizations.

In this work we have studied the exit problem for small noise diffusions. In particular, we have shown the existence of possible asymmetries in the case in which the flow generated by the drift admits a saddle point. The proof is restricted to the 2-dimensional setting. Let us discuss about this particular restriction.

Our method of proof was to transform the original equation into a very specific non-linear equation known as normal form. Then, we obtained several estimates that intensively uses the smallness of the noise and the specific form of the nonlinearity in the normal form. Let us recall the form of the nonlinearity.

In our case, the nonlinearity in the normal form is given by a finite sum of resonant monomials (see Section 2.3) of the form $(x_1^{\alpha_1^+} x_2^{\alpha_2^+}, x_1^{\alpha_1^-} x_2^{\alpha_2^-})$, where $(\alpha_1^\pm, \alpha_2^\pm) \in \mathbb{Z}^2$ satisfy the resonance relations

$$\alpha_1^\pm \lambda_+ - \alpha_2^\pm \lambda_- = \pm\lambda_\pm,$$

of some order $r = \alpha_1^\pm + \alpha_2^\pm \geq 2$. If we were to generalize the argument in Chapter 2 to the d-dimensional case, we would need to take into account the particular form that the nonlinearity would have in the normal form. Indeed, there are two points to consider:

1. The resonant monomials of order $r \geq 2$ are of the form

$$(x_1^{\alpha_{1,1}} \cdots x_d^{\alpha_{1,d}}, ..., x_1^{\alpha_{d,1}} \cdots x_d^{\alpha_{d,d}}),$$

where the vector $\alpha_i = (\alpha_{i,1}, ..., \alpha_{i,d}) \in \mathbb{Z}^d$ satisfies

$$\begin{aligned} \alpha_{i,1}\lambda_1 + ... + \alpha_{i,d}\lambda_d &= \lambda_i \\ \alpha_{i,1} + ... + \alpha_{i,d} &= r, \end{aligned}$$

for each $i = 1, ..., d$. Here $\lambda_1, ..., \lambda_d$ are the eigenvalues of the matrix $\nabla b(0)$.

2. According to [38, Theorem 3,Section 2], the nonlinearity N, after being transformed by a normal form transformation of degree $R > 1$, will be of the form
$$N(x) = P(x) + Q(x),$$
where P is a finite sum of resonant monomials, and Q is a correction of order $|x|^{R+1}$ (as $|x| \to 0$) when the vector of eigenvalues $\lambda = (\lambda_1, ..., \lambda_d)$ is not one-resonant and identically 0 when λ is one-resonant.

The first point implies that to obtain the exponents $\alpha_{i,j}$ more combinatorial work than the one put in Section 2.3.2 is needed. Still this is not the biggest difficulty. The biggest difficulty relies on the lack of structure of the correction Q in the case λ is not one-resonant. Indeed, to have any hope that our techniques in Chapter 2 work, we require at least that whenever $\alpha_{i,k} \neq 0$ for some $k < i$, then $\alpha_{i,j} \neq 0$ for some $k < j \leq d$ (this is in the case we order the eigenvalues as usual: $\operatorname{Re}\lambda_1 \geq ... \geq \operatorname{Re}\lambda_d$). There is no guarantee that a condition of this form holds in the not one-resonant case. In conclusion, a higher dimensional analogue for the saddle case can be obtained using the techniques presented in this theses only in the one-resonant case. This is so, unless the particular structure that the eigenvalues $\lambda_1, ..., \lambda_d$ have in this case implies that a normal form transformation can be chosen so that the non-linearity of the transformed drift is a finite sum of resonant monomials with no correction. As far as the author knows, this is an unsolved issue in normal form theory.

There are still some results to be filled in order to complete the case in which the deterministic flow has a saddle, and hence the case in which it admits an heteroclinic network. This result is worthwhile pursuing since the implications of the asymmetry found in Chapter 2 has very interesting analogues in higher dimensions as famous chaotic systems (such as the Lorentz system) in higher dimensions exhibit homoclinic behavior (see [64, Chapters 27,30 and 31] for further examples).

4.2.1 A non-smooth transformation alternative

As discussed in Section 2.3.2, it is possible to conjugate a nonlinear equation to a linear one. The restriction for Itô equations is that this transformation has to be at least C^2. Recent results have extended Itô's formula for functions with less smoothness. The first result of this nature is the well known Tanaka's formula [54, Chapter IV], which relies on the existence of local time for one dimensional semimartingales to extend the range of applicability of Itô's formula to convex functions. For higher dimensional semimartingales there is no local time, so there was no immediate high dimensional analogue for Tanaka's formula. For a long time Tanaka's formula remain the more general change of variables (in terms of smoothness requirements) known. Recent studies have established change of variables for higher dimensional semimartingales with less smoothness [57], [58], [29], [26]. Let us give a brief (and informal) comment about this setting.

Consider $f : \mathbb{R}^d \to \mathbb{R}$ to be a continuously differentiable function. Let Z be a semimartingale in $\mathbb{R}^d$ with $Z(t) = V(t) + M(t)$, M being a martingale and V a stochastic process with bounded variation paths. Then [57], [58], [29], [26] agree that if the quadratic covariation $[f(Z), Z^j]$ is well defined for every $1 \leq j \leq d$, then Itô's formula holds:

$$f(Z(t)) = f(Z(0)) + \int_0^t \nabla f(Z(s))dZ(s) + \frac{1}{2}\sum_{i,j=1}^{d}[\partial_{x_j} f(Z), Z^j](t).$$

We recall the definition of quadratic covariation (see [54, Section V.5]):

Definition 56 *Let H and J be two continuous stochastic processes in $\mathbb{R}$. The quadratic covariation of H and J, denoted as $[H, J]$, is, when it exist, the continuous process of finite variation over compacts, such that for any sequence σ_n of random partitions tending to the identity,*

$$[H, J] = H(0)J(0) + \lim_{n\to\infty}\sum_i (H^{T^n_{i+1}} - H^{T^n_i})(H^{T^n_{i+1}} - H^{T^n_i}), \tag{4.1}$$

uniformly over compacts in probability. Here, for any random $S > 0$, the process H^S is short for $t \mapsto H(t \wedge S)$, and σ_n is the sequence $0 = T^n_0 \leq \ldots \leq T^n_{k_n}$, where $\sup_i(T^n_{i+1} - T^n_i) \to 0$, $k_n \to \infty$, and, $T^n_{k_n} \to \infty$ as $n \to \infty$.

For our diffusion process X_ϵ, there are several problems to consider. One is to show that $[\partial_{x_j} f(X_\epsilon), X^i_\epsilon]$ is well defined. The other, is to prove that

$$\epsilon^{-1}[\partial_{x_j} f(X_\epsilon), X^i_\epsilon] \xrightarrow{\mathbf{P}} 0, \quad \epsilon \to 0. \tag{4.2}$$

Once this is established, the result in Chapter 2 follows immediately.

In order to show that $[\partial_{x_j} f(X_\epsilon), X_\epsilon^i]$ is well defined, the proposal in [57], [58], [29] is to use the theory of reversible diffusions proposed in [50]. Indeed, assume for a moment that we know that, for a fixed $T > 0$, $\hat{X}_\epsilon(t) = X_\epsilon(T - t)$ is a diffusion. Then, observe that (4.1) for $H = \partial_{x_j} f(X_\epsilon)$ and $J = X_\epsilon^i$ can be written as

$$[\partial_{x_j} f(X_\epsilon), X_\epsilon^i](t) = -\int_0^t \partial_{x_j} f(X_\epsilon(s)) dX_\epsilon^i(s) - \int_{T-t}^T \partial_{x_j} f(\hat{X}(s)) d\hat{X}_\epsilon^i(s), \tag{4.3}$$

where both integrals are Itô integrals with respect to different filtrations. In order to use this formula to prove (4.2) the first attempt may be to get the generator of $\hat{X}_\epsilon$. Under several assumptions (the most important one being the ellipticity of the noise) in [50] it is proved that, for a fixed time $T > 0$, $X(T - \cdot)$ is also a diffusion with the same diffusion matrix and with drift $\hat{b} = -b(x) + \nabla \log p_{T-t,T}(x, X_\epsilon(T))$, where $p_{T-t,T}$ is the transition density of the Markov process X_ϵ (which existence is proved in [50]). Hence in order to establish (4.2) we first would need to have a bound in ϵ of $\nabla \log p_{T-t,T}$. This quantity is of interest in control theory [28], but there is, as far as the author knows,no reference to an estimate in $\epsilon > 0$. Another option that avoids this estimate is to extend the filtration $\mathcal{F}^W$ to the minimum complete filtration that includes $\mathcal{F}^W$ such that $X_\epsilon(T)$ is measurable and write Doob-Meyer decomposition for the process X_ϵ with respect to this filtration.

This is still undergoing work, that is promising not only because it allows to prove the results included in this thesis, but also because it uses several tools of modern stochastic analysis.

4.3 Scaling limits

In this thesis we proved a scaling limit for the exit problem for two cases, the case in which the flow S has a unique saddle and the Levinson case. The idea will be to prove scaling limits for more general systems. In particular, recall that if the quasipotential has a unique minimizer z, then the exit point $X_\epsilon(\tau_\epsilon)$ converges to it in probability as $\epsilon \to 0$. By a scaling limit, we mean find an $\alpha > 0$ such that the distribution of $\epsilon^{-\alpha}(z - X_\epsilon(\tau_\epsilon^D))$ is tight.

Let $V : D \times \partial D \to [0, \infty)$ be the quasipotential given by (1.7):

$$V(x, y) = \inf_{T>0} \{I_T^x(\varphi) : \varphi(T) = y, \varphi([0, T]) \subset D \cup \partial D\}$$

In order to state the claim, given $x_0 \in D$, let $\mathcal{M}_{x_0} \subset \partial D$ be the set of minimizers of $y \mapsto V(x_0, y)$. The claim is the following:

Claim 57 *Suppose $\mathcal{M}_{x_0}$ is finite $\mathcal{M}_{x_0} = \{e_1, ..., e_q\}$. There is a probability distribution ν over $\mathcal{M}$, a random number $\alpha \in (0, 1]$ and a family of random variables $(\xi_\epsilon)_{\epsilon>0}$ such that the exit can be written as*

$$X_\epsilon(\tau_\epsilon) = \nu_1 e_1 + ... + \nu_q e_q + \epsilon^\alpha \xi_\epsilon.$$

Further, there is a random variable ξ_0 so that $\xi_\epsilon \to \xi_0$ in distribution as $\epsilon \to 0$.

The results of this thesis imply this claim in the case the flow S admits an heteroclinic network (see Section 1.4.2). The proof was done by solving two simple cases (saddle point and Levinson case) and then using a Poincaré distributional map argument for each critical point in the network.

Here we shall proceed similarly: start from simple cases with random initial conditions so that a Poincaré argument can be applied. The proposal is to choose as the base case the well developed stable case [34]: $0 \in D$ and D is contained in the basin of attraction of 0. It is known that if the domain D is attracted to the origin and $M_{x_0} = \{e\}$ then $X_\epsilon(\tau_\epsilon) \to e$ in probability. Moreover, if there is a unique extreme trajectory φ_0 (the one that realizes the minimum in V) then for every $\delta > 0$,

$$\lim_{\epsilon \to 0} \mathbf{P}_{x_0} \left\{ \sup_{\theta_\epsilon \le t \le \tau_\epsilon} |X_\epsilon(t) - \varphi_0(t - \theta_\epsilon + \theta_0)| < \delta \right\} = 0,$$

where θ_ϵ (θ_0) is the last time X_ϵ (φ_0) hits a ball of arbitrary small (but fixed) radius around the origin. From the perspective introduced in Chapter 3, consider the process conditioned on exit close to e. This process is a semimartingale with the same diffusion matrix as the original process, but with a drift of the form $b_\epsilon = -b(x) + \epsilon^2 \varphi_\epsilon(t, x)$, where φ_ϵ is uniformly bounded. Hence, our results in the Levinson case of Chapter 3 apply.

Once this result is established, we can follow the same pattern as in [34] to study possible asymmetric behavior in metastable process. With this development we can show that the idea of random Poincaré maps apply to a general dynamical system.

Appendix A

Large Deviations

Large deviation theory is a mixture of probability theory, analysis, variational calculus, point set topology among others. This theory has been used for different purposes. In Section 1.1 we discussed the role played by large deviation theory in the development of Freidlin-Wentzell theory. The purpose of this chapter is to provide a quick reference to large deviation theory as needed to understand Section 1.1.

We present the general theory of large deviations. The theory was first formulated in the right degree of abstraction by Varadhan [61], we follow [52] in this exposition. In Section A.1 we begin with the basic definitions. In Section A.2 we present the large deviation results related to diffusion processes.

A.1 Large Deviations Principle (LDP)

Let $\mathcal{X}$ be a Polish metric space with metric function $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$. By a probability measure on $\mathcal{X}$, we mean a probability measure on the Borel sigma algebra on $\mathcal{X}$. We will give the general definition of large deviation principle for a family of probability measures on $\mathcal{X}$. First, recall the following definition.

Definition 58 *The function $f : \mathcal{X} \to [-\infty, \infty]$ is lower semi-continuous if it satisfies any of the following equivalent properties:*

1. $\liminf_{n\to\infty} f(x_n) \geq f(x)$ *for all sequences $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ and all points $x \in \mathcal{X}$ such that $x_n \to x$ in $\mathcal{X}$.*

2. *For all $x \in \mathcal{X}$, $\lim_{\delta\to 0} \inf_{y\in B_\delta(x)} f(y) = f(x)$, where $B_\delta(x) = \{y \in \mathcal{X} : d(x, y) < \delta\}$.*

3. *f has closed level sets, that is, $f^{-1}([-\infty, c]) = \{x \in \mathcal{X} : f(x) \leq c\}$ is closed for all $c \in \mathbb{R}$.*

Here are the key definitions of large deviation theory:

Definition 59 *The function $I : \mathcal{X} \to [0, \infty]$ is called a rate function if*

1. *$I \not\equiv \infty$,*

2. *I is lower semi-continuous,*

3. *I has compact level sets.*

Definition 60 *A family of probability measures $(\mathbf{P})_{\epsilon>0}$ on $\mathcal{X}$ is said to satisfy , as $\epsilon \to 0$,the large deviation principle (LDP) with rate $\alpha_\epsilon \to 0$ and rate function I if*

1. *I is a rate function,*

2. *$\limsup_{\epsilon\to 0} \alpha_\epsilon \log \mathbf{P}_\epsilon(C) \leq -I(C)$, for every $C \subset \mathcal{X}$ closed,*

3. *$\liminf_{\epsilon\to 0} \alpha_\epsilon \log \mathbf{P}_\epsilon(O) \geq -I(O)$, for every $O \subset \mathcal{X}$ open.*

Here the bounds are in terms of the set function defined by

$$I(S) = \inf_{s\in S} I(x), \quad S \subset \mathcal{X}.$$

The goal of large deviation theory is to build up an arsenal of theorems based on these two definitions. We will not describe most of this theorems, since they are out of the scope for the present text. The interested reader is invited to consult the standard monographs on the subject [23, Chapter 4], [24, Chapter III], [52, Chapter 2]. The only theorem that we cite is the so called contraction principle. First, we give some remarks

Remark 6 1. *It is a standard exercise to show that once the large deviation principle is satisfied, the rate function I is unique.*

2. *In Definition 60 it is crucial to make a difference between open and closed sets. Naively, one might try to replace the second and third conditions with the stronger requirement that*

$$\lim_{\epsilon\to 0} \alpha_\epsilon \mathbf{P}_\epsilon(S) = -I(S), \quad S \subset \mathcal{X}.$$

However, there are examples that show that this would be far too restrictive.

We now present the contraction principle:

Theorem 61 *Let $(\mathbf{P})_{\epsilon>0}$ be a family of probability measures on $\mathcal{X}$ that satisfies the LDP, as $\epsilon \to 0$, with rate function α_ϵ and with rate function I. Let $\mathcal{Y}$ be a Polish space, $T : \mathcal{X} \to \mathcal{Y}$ a continuous map, and $\mathbb{Q}_\epsilon = \mathbf{P}_\epsilon \circ T^{-1}$ an image probability measure. Then, the family $(\mathbb{Q}_\epsilon)_{\epsilon>0}$ satisfies the LDP on $\mathcal{Y}$ with rate α_ϵ and with rate function J given by*

$$J(y) = \inf_{x \in \mathcal{X}: T(x)=y} I(x),$$

with the convention $\inf_\emptyset I = \infty$.

A.2 Freidlin-Wentzell LDP

In this section we present the large deviation results that Freidlin-Wentzell theory is based on.

Given $T > 0$, let $W(t), t \in [0,T]$, be a standard Brownian motion in R^d. Consider the process $W_\epsilon(t) = \epsilon W(t)$, and let $\mathbf{P}_\epsilon^W$ be the probability measure induced by W_ϵ on $C([0,T];\mathbb{R}^d)$, the space of all continuous functions $\varphi : [0,T] \to \mathbb{R}^d$ equipped with the supremum norm topology. We first state the LDP for W_ϵ derived by Schilder [59]:

Theorem 62 *The family of probability measures $(\mathbf{P}_\epsilon^W)_{\epsilon>0}$ on $C([0,T];\mathbb{R}^d)$ satisfy a LDP with rate ϵ^2 and with rate function*

$$J_T(\phi) = \begin{cases} \frac{1}{2}\int_0^T |\dot{\phi}(s)|^2 ds & , \quad \phi \in H^1 \\ \infty & , \quad \textit{otherwise} \end{cases}$$

Here H^1 is the space of absolutely continuous functions with square integrable derivative.

The simple case in which the process X_ϵ is the strong solution of

$$dX_\epsilon(t) = b(X_\epsilon(t))dt + \epsilon dW(t)$$

is a consequence of Theorem 62 and the Contraction Principle 61. Indeed, let $F : C([0,T];\mathbb{R}^d) \to C([0,T];\mathbb{R}^d)$ be the map defined by $f = F(g)$, where f is the unique solution of

$$f(t) = \int_0^T b(f(s))ds + g(t).$$

Then, after noticing that F is continuous and some calculation, Theorem 62 implies the following result:

Corollary 63 *The law of X_ϵ on $C([0,T];\mathbb{R}^d)$ satisfies a LDP with rate ϵ^2 and with rate function*

$$J'_T = \begin{cases} \frac{1}{2}\int_0^T |\dot{\phi}(s) - b(s)|^2 ds & , \quad \phi \in H^1 \\ \infty & , \quad \mathit{otherwise} \end{cases}.$$

Now, consider X_ϵ to be the solution of our typical SDE

$$dX_\epsilon(t) = b(X_\epsilon(t))dt + \epsilon\sigma(X_\epsilon(t))dW(t).$$

As said on Section 1.1 a LDP for this process is the base of Freidlin-Wentzell theory. It turns out that to obtain a LDP for the law of X_ϵ the contraction principle does not apply. Instead, raw approximations have to be made. We state the theorem without a proof (see [23, Section 4.2] or [34, chapter 3] for a proof).

Theorem 64 (Freidlin-Wentzell [34]) *Let $H^1_{0,T}$ be the space of all absolutely continuous functions from $[0,T]$ to $\mathbb{R}^d$ with square integrable derivatives. Define the functional I^x_T by*

$$I^x_T(\varphi) = \frac{1}{2}\int_0^T \langle \dot{\varphi}(s) - b(\varphi(s)), a^{-1}(\varphi(s))(\dot{\varphi}(s) - b(\varphi(s)))\rangle ds, \tag{A.1}$$

if $\varphi \in H^1_{0,T}$ and $\varphi(0) = x$, and ∞ otherwise. Here b is the drift in (1.1) *and $a = \sigma^T\sigma$, with σ the diffusion matrix in* (1.1).

Then for each $x \in \mathbb{R}^d$ and $T > 0$ the family $(\mathbf{P}^\epsilon_x)_{\epsilon>0}$ satisfies a Large Deviation Principle on $C([0,T];\mathbb{R}^d)$ equipped with uniform norm at rate ϵ^2 with good rate function I^x_T.

Appendix B

Appendix to Section 1.2.1

The purpose of this appendix is to present any technical material left out in Chapter 1 Section 1.2.1. This appendix (in contrast with Appendix A) contains original material in the simple case discussed on Section B. Let us recall the setting from that section.

Given two positive numbers $\lambda_{\pm} > 0$, consider the diffusion $X_\epsilon = (x_\epsilon^1, x_\epsilon^2)$

$$dX_\epsilon(t) = \mathrm{diag}(\lambda_+, -\lambda_-)X_\epsilon(t)dt + \epsilon dW(t).$$

Let $\delta > 0$ and $D = (-\delta, \delta) \times (-\delta, \delta) \subset \mathbb{R}^2$. We study the exit problem of X_ϵ from D. We start the diffusion X_ϵ inside D: $X_\epsilon(0) = (0, x_0) \in D$.

Recall that in Section 1.2.1, we used Itô's formula in each coordinate to write Duhamel principle for x_ϵ^1 and x_ϵ^2. Here we rewrite identities (1.8) and (1.9) for easier reference:

$$x_\epsilon^1(t) = \epsilon e^{\lambda_+ t} \int_0^t e^{-\lambda_+ s} dW_1(s), \tag{B.1}$$

$$x_\epsilon^2(t) = e^{-\lambda_- t} x_0 + \epsilon \int_0^t e^{-\lambda_-(t-s)} dW_2(s). \tag{B.2}$$

Let $\mathcal{N}(t)$ denote the stochastic integral in (B.1).

Recall that τ_ϵ^δ is defined as

$$\tau_\epsilon^\delta = \inf\left\{t > 0 : |x_\epsilon^1(t)| \geq \delta\right\}.$$

First we prove that τ_ϵ^δ is finite with probability 1. This is a general fact that can be found in the literature, for example in [8, Proposition 1.8.2], but we chose to prove it directly from Duhamel principle. We do this in order to stress the importance of such a representation in our setting. Without any further discussion, we go into the results.

Lemma 65 *For every $\delta > 0$ and $\epsilon > 0$, $\tau_\epsilon^\delta < \infty$ $\mathbf{P}-$a.s.*

Proof. Let $n \in \mathbb{N}$, it is enough to show that $\mathbf{P}\{\tau_\epsilon^\delta > n\} \to 0$ as $n \to \infty$. Observe that (B.1) implies that

$$\begin{aligned}\mathbf{P}\{\tau_\epsilon^\delta > n\} &= \mathbf{P}\left\{\sup_{t\in[0,n]} \epsilon e^{\lambda_+ t}|\mathcal{N}(t)| < \delta\right\}\\ &\leq \mathbf{P}\left\{\sup_{t\in[n/2,n]} \epsilon e^{\lambda_+ t}|\mathcal{N}(t)| < \delta\right\}\\ &\leq \mathbf{P}\left\{\epsilon e^{\lambda_+ n/2}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| < \delta\right\}.\end{aligned}$$

Here the last two inequalities follow from the properties of the supremum and the exponential function respectively. Take $n_0 \in \mathbb{N}$ such that $\epsilon^{-1}\delta < e^{\lambda_+ n/4}$, for every $n \geq n_0$. Then, for every $n \geq n_0$,

$$\begin{aligned}\mathbf{P}\left\{\epsilon e^{\lambda_+ n/2}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| < \delta\right\} &\leq \mathbf{P}\left\{\epsilon e^{\lambda_+ n/2}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| < \delta, e^{\lambda_+ n/4}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| \geq 1\right\}\\ &\quad + \mathbf{P}\left\{e^{\lambda_+ n/4}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| < 1\right\}\\ &\leq \mathbf{P}\left\{e^{\lambda_+ n/4}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| < 1\right\}.\end{aligned}$$

The proof will be finished as soon as we can show that the last probability converges to 0. To see this, note that, for every $t > 0$, the random variable $\mathcal{N}(t)$ is a zero mean gaussian random variable with variance $\int_0^t e^{-2\lambda_+ t} ds = \frac{1-e^{-2\lambda_+ t}}{2\lambda_+}$. Denote

$$\alpha_n = e^{-\lambda_+ n/4}\sqrt{\frac{2\lambda_+}{1-e^{-2\lambda_+ n}}}.$$

The result follows since, $\alpha_n \to 0$, as $n \to \infty$, and

$$\begin{aligned}\mathbf{P}\left\{e^{\lambda_+ n/4}\sup_{t\in[n/2,n]} |\mathcal{N}(t)| < 1\right\} &\leq \mathbf{P}\left\{e^{\lambda_+ n/4}|\mathcal{N}(n)| < 1\right\}\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\alpha_n}^{\alpha_n} e^{-r^2/2} dr.\end{aligned}$$

■

Lemma 66 *For any $\delta > 0$,*

$$\tau_\epsilon^\delta \xrightarrow{\mathbf{P}} \infty,$$

as $\epsilon \to 0$.

Proof. Let $\delta > 0$. It is enough to prove that $\mathbf{P}\{\tau_\epsilon^\delta > T\} \to 0$, $\epsilon \to 0$, for any $T > 0$. Use Duhamel principle (B.1) to get

$$\begin{aligned}\mathbf{P}\{\tau_\epsilon^\delta > T\} &= \mathbf{P}\left\{\sup_{t\le T\wedge\tau_\epsilon^\delta} |x_\epsilon^1(t)| > \delta\right\}\\ &\le \mathbf{P}\left\{\epsilon e^{\lambda_+ T}\sup_{t\le T\wedge\tau_\epsilon^\delta}\left|\int_0^t e^{-\lambda_+ s}ds\right| > \delta\right\}.\end{aligned}$$

The last inequality, Chebyshev inequality [21, Section 3.2], BDG inequality [41, Proposition 3.3.28] and Itô isometry [41, Proposition 2.10] imply that for some constant C_1,

$$\begin{aligned}\mathbf{P}\{\tau_\epsilon^\delta > T\} &\le C_1 e^{2\lambda_+ T}\epsilon^2 \mathbf{E}\int_0^{T\wedge\tau_\epsilon^\delta} e^{-2\lambda_+ s}dW(s)\\ &\le \frac{C_1}{\lambda_+}e^{2\lambda_+ T}\epsilon^2.\end{aligned}$$

This proves our result. ■

The last technical step in this appendix is about the convergence of the random variable $\mathcal{N}_\epsilon$. Recall that $\mathcal{N}_\epsilon = \mathcal{N}(\tau_\epsilon^\delta)$ and

$$\mathcal{N} = \int_0^\infty e^{-\lambda_+ s}dW(s).$$

Lemma 67 *As $\epsilon \to 0$, $\mathcal{N}_\epsilon \to \mathcal{N}$ in probability.*

Proof. The lemma is a consequence of Itô isometry and Lemma 66. Let $\gamma > 0$ and $T_\gamma = -(2\lambda_+)^{-1}\log(\gamma\lambda_+) > 0$. Due to Lemma 66 we can find $\epsilon_0 > 0$ such that

$$\mathbf{P}\{\tau_\epsilon^\delta > T_\gamma\} \le \gamma\lambda_+, \tag{B.3}$$

for every $\epsilon \in (0, \epsilon_0)$. Use Itô isometry and (B.3) to obtain

$$\begin{aligned}\mathbf{E}|\mathcal{N}_\epsilon - \mathcal{N}|^2 &= \mathbf{E}\int_{\tau_\epsilon^\delta}^\infty e^{-2\lambda_+ s}ds\\ &\le \frac{e^{-2\lambda_+ T_\gamma}}{2\lambda_+} + \frac{1}{2\lambda_+}\mathbf{P}\{\tau_\epsilon^\delta < T_\gamma\}\\ &\le \gamma,\end{aligned}$$

for every $\epsilon \in (0, \epsilon_0)$. The result follows since $\gamma > 0$ is arbitrary and L^2 convergence implies convergence in probability. ■

Bibliography

[1] Sergio Angel Almada Monter and Yuri Bakhtin. Normal forms approach to diffusion near hyperbolic equilibria. *Submitted to Nonlinearity; also available at http://arxiv.org/abs/1006.3000.*

[2] Sergio Angel Almada Monter and Yuri Bakhtin. Scaling limit for the diffusion exit problem in the Levinson case. *Submitted to Stoch. Process. Appl.; also available at http://arxiv.org/abs/1006.2766.*

[3] Ludwig Arnold and Peter Imkeller. Normal forms for stochastic differential equations. *Probability Theory and Related Fields*, 110:559–588, 1998. 10.1007/s004400050159.

[4] Yuri Bakhtin. Noisy heteroclinic networks. *Probability Theory and Related Fields, in print; also available at http://arxiv.org/abs/0712.3952.*

[5] Yuri Bakhtin. Exit asymptotics for small diffusion about an unstable equilibrium. *Stochastic Process. Appl.*, 118(5):839–851, 2008.

[6] Yuri Bakhtin. Small noise limit for diffusions near heteroclinic networks. *Dynamical Systems: An International Journal*, 25:413–431, 2010.

[7] Yuri Bakhtin. Small noise limit for diffusions near heteroclinic networks. *Dynamical Systems, in print*, 2010.

[8] Richard F. Bass. *Diffusions and elliptic operators*. Probability and its Applications (New York). Springer-Verlag, New York, 1998.

[9] Gérard Ben Arous and Fabienne Castell. Flow decomposition and large deviations. *J. Funct. Anal.*, 140(1):23–67, 1996.

[10] Roberto Benzi, Giorgio Parisi, Alfonso Sutera, and Angelo Vulpiani. A theory of stochastic resonance in climatic change. *SIAM J. Appl. Math.*, 43(3):565–478, 1983.

[11] Nils Berglund and Barbara Gentz. Metastability in simple climate models: pathwise analysis of slowly driven Langevin equations. *Stoch. Dyn.*, 2(3):327–356, 2002. Special issue on stochastic climate models.

[12] Nils Berglund and Barbara Gentz. Pathwise description of dynamic pitchfork bifurcations with additive noise. *Probab. Theory Related Fields*, 122(3):341–388, 2002.

[13] Nils Berglund and Barbara Gentz. *Noise-induced phenomena in slow-fast dynamical systems*. Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2006. A sample-paths approach.

[14] Nils Berglund and Barbara Gentz. Anomalous behavior of the Kramers rate at bifurcations in classical field theories. *J. Phys. A*, 42(5):052001, 9, 2009.

[15] Ju. N. Blagoveščenskiĭ. Diffusion processes depending on a small parameter. *Teor. Verojatnost. i Primenen.*, 7:135–152, 1962.

[16] B.-Z. Bobrovsky, M.M. Zakai, and O. Zeitouni. Error bounds for the nonlinear filtering of signals with small diffusion coefficients. *Information Theory, IEEE Transactions on*, 34(4):710 –721, July 1988.

[17] Caroline Cardon-Weber. Large deviations for a Burgers'-type SPDE. *Stochastic Process. Appl.*, 84(1):53–70, 1999.

[18] Fabienne Castell. Asymptotic expansion of stochastic flows. *Probab. Theory Related Fields*, 96(2):225–239, 1993.

[19] Sandra Cerrai and Michael Röckner. Large deviations for stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term. *Ann. Probab.*, 32(1B):1100–1139, 2004.

[20] Shui-Nee Chow, Cheng Zhi Li, and Duo Wang. *Normal forms and bifurcation of planar vector fields*. Cambridge University Press, Cambridge, 1994.

[21] Kai Lai Chung. *A course in probability theory*. Academic Press Inc., San Diego, CA, third edition, 2001.

[22] Martin V. Day. On the exit law from saddle points. *Stochastic Process. Appl.*, 60(2):287–311, 1995.

[23] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[24] Frank den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[25] Allen Devinatz and Avner Friedman. The asymptotic behavior of the solution of a singularly perturbed Dirichlet problem. *Indiana Univ. Math. J.*, 27(3):527–537, 1978.

[26] Nathalie Eisenbaum. Local time-space stochastic calculus for Lévy processes. *Stochastic Process. Appl.*, 116(5):757–778, 2006.

[27] Jin Feng, Martin Forde, and Jean-Pierre Fouque. Short-maturity asymptotics for a fast mean-reverting Heston stochastic volatility model. *SIAM J. Financial Math.*, 1:126–141, 2010.

[28] Wendell H. Fleming and H. Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Springer, New York, second edition, 2006.

[29] Hans Föllmer and Philip Protter. On Itô's formula for multidimensional Brownian motion. *Probab. Theory Related Fields*, 116(1):1–20, 2000.

[30] Eric Forgoston, Lora Billings, and Ira B. Schwartz. Accurate noise projection for reduced stochastic epidemic models. *Chaos*, 19(4):043110, 15, 2009.

[31] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. *Derivatives in financial markets with stochastic volatility*. Cambridge University Press, Cambridge, 2000.

[32] M. Freidlin and L. Koralov. Metastability for nonlinear random perturbations of dynamical systems. *Stochastic Process. Appl.*, 120(7):1194–1214, 2010.

[33] M. Freidlin and L. Koralov. Nonlinear stochastic perturbations of dynamical systems and quasi-linear parabolic PDE's with a small parameter. *Probab. Theory Related Fields*, 147(1-2):273–301, 2010.

[34] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen*

Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, New York, second edition, 1998. Translated from the 1979 Russian original by Joseph Szücs.

[35] Mark Freidlin. *Markov processes and differential equations: asymptotic problems*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1996.

[36] Peter Hanggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Modern Phys.*, 62(2):251–341, 1990.

[37] Philip Hartman. *Ordinary differential equations*. John Wiley & Sons Inc., New York, 1964.

[38] Yu. S. Il′yashenko and S. Yu. Yakovenko. Finitely smooth normal forms of local families of diffeomorphisms and vector fields. *Uspekhi Mat. Nauk*, 46(1(277)):3–39, 240, 1991.

[39] S. Kamin. Elliptic perturbation of a first-order operator with a singular point of attracting type. *Indiana Univ. Math. J.*, 27(6):935–952, 1978.

[40] Shoshana Kamin. On elliptic equations with a small parameter in the highest derivative. *Comm. Partial Differential Equations*, 4(6):573–593, 1979.

[41] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988.

[42] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[43] Ju. I. Kifer. Some theorems on small random perturbations of dynamical systems. *Uspehi Mat. Nauk*, 29(3(177)):205–206, 1974.

[44] Yuri Kifer. The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point. *Israel J. Math.*, 40(1):74–96, 1981.

[45] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.

[46] Norman Levinson. The first boundary value problem for $\varepsilon\Delta u + A(x,y)u_x + B(x,y)u_y + C(x,y)u = D(x,y)$ for small ε. *Ann. of Math. (2)*, 51:428–445, 1950.

[47] Norman Levinson. *Selected papers of Norman Levinson. Vol. 2.* Contemporary Mathematicians. Birkhäuser Boston Inc., Boston, MA, 1998. Edited by John A. Nohel and David H. Sattinger.

[48] K. R. Meyer. Counterexamples in dynamical systems via normal form theory. *SIAM Rev.*, 28(1):41–51, 1986.

[49] Toshio Mikami. Limit theorems on the exit problems for small random perturbations of dynamical systems. II. *Kodai Math. J.*, 17(1):48–68, 1994.

[50] A. Millet, D. Nualart, and M. Sanz. Integration by parts and time reversal for diffusion processes. *Ann. Probab.*, 17(1):208–238, 1989.

[51] A. Millet, D. Nualart, and M. Sanz. Large deviations for a class of anticipating stochastic differential equations. *Ann. Probab.*, 20(4):1902–1931, 1992.

[52] Enzo Olivieri and Maria Eulália Vares. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge, 2005.

[53] Lawrence Perko. *Differential equations and dynamical systems*, volume 7 of *Texts in Applied Mathematics.* Springer-Verlag, New York, third edition, 2001.

[54] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[55] Mikhail I. Rabinovich, Ramon Huerta, and Valentin Afraimovich. Dynamics of sequential decision making. *Physical Review Letters*, 97(18):188103, 2006.

[56] Mikhail I. Rabinovich, Ramón Huerta, Pablo Varona, and Valentin S. Afraimovich. Transient cognitive dynamics, metastability, and decision making. *PLoS Comput. Biol.*, 4(5):e1000072, 9, 2008.

[57] F. Russo and P. Vallois. Itô formula for C^1-functions of semimartingales. *Probab. Theory Related Fields*, 104(1):27–41, 1996.

[58] Francesco Russo and Pierre Vallois. Elements of stochastic calculus via regularization. In *Séminaire de Probabilités XL*, volume 1899 of *Lecture Notes in Math.*, pages 147–185. Springer, Berlin, 2007.

[59] M. Schilder. Some asymptotic formulas for Wiener integrals. *Trans. Amer. Math. Soc.*, 125:63–85, 1966.

[60] Zeev Schuss and Bernard J. Matkowsky. The exit problem: a new approach to diffusion across potential barriers. *SIAM J. Appl. Math.*, 36(3):604–623, 1979.

[61] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, 19:261–286, 1966.

[62] A. D. Ventcel′ and M. I. Freĭdlin. Small random perturbations of dynamical systems. *Uspehi Mat. Nauk*, 25(1 (151)):3–55, 1970.

[63] A. D. Ventcel′ and M. I. Freĭdlin. Certain problems that concern stability under small random perturbations. *Teor. Verojatnost. i Primenen.*, 17:281–295, 1972.

[64] Stephen Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*, volume 2 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2003.

[65] O. Zeitouni. On the filtering of noise-contaminated signals observed via hard limiters. *Information Theory, IEEE Transactions on*, 34(5):1041 –1048, September 1988.

[66] Ofer Zeitouni and Moshe Zakai. On the optimal tracking problem. *SIAM J. Control Optim.*, 30(2):426–439, 1992.